



QY 241 AAGGAACACAGGCTCAGAGAACAGATGGAGGAGAGCAGAAAGCAATAGAGAAATC 300  
Db 253 AAGGAACACAGGCTCAGAGAACAGATGGAGGAGAGCAGAAAGCAATAGAGAAATC 312  
QY 301 CTTGGGAAAAACAACGGAACCTACTAAGAAGAGTGAGAAGCAGTTACAGTGTTCGTA 360  
Db 313 CTTGGGAAAAACAACGGAACCTACTAAGAAGAGTGAGAAGCAGTTACAGTGTTCGTA 372  
QY 361 ACACACTGCATTAAAGTGGCAATGCTTGCCTGGACCCAGCCCTGCAGCCCAACATCACC 420  
Db 373 ACACACTGCATTAAAGTGGCAATGCTTGCCTGGACCCAGCCCTGCAGCCCAACATCACC 432  
QY 421 CTGAAGGACCTGCATCTCTTTATCCTCTTTTCATCTCGCAGTGACATTTCAATGTT 480  
Db 433 CTGAAGGACCTGCATCTCTTTATCCTCTTTTCATCTCGCAGTGACATTTCAATGTT 492  
QY 481 GCAAAACCAAAAAACCTCTCAACCAATGCTCAGTTGTTGTTTGGACAGTACTAAGAT 540  
Db 493 GCAAAACCAAAAAACCTCTCAACCAATGCTCAGTTGTTGTTTGGACAGTACTAAGAT 552  
QY 541 GTTGAAGATGCCACCTCTGACTGCTTTAAAGGAAATPAGCAGACAGATGGAGGGC 600  
Db 553 GTTGAAGATGCCACCTCTGACTGCTTTAAAGGAAATPAGCAGACAGATGGAGGGC 612  
QY 601 TACTTGACACAGATAAAGATCCCTGATAGTGTCTAATGCAAGATTTGGCTTTCCTG 660  
Db 613 TACTTGACACAGATAAAGATCCCTGATAGTGTCTAATGCAAGATTTGGCTTTCCTG 672  
QY 661 AGTGGATTTCCACCAACATTCAGGAAACAAATCAACTAAAACAAAATGGCCAGAAAT 720  
Db 673 AGTGGATTTCCACCAACATTCAGGAAACAAATCAACTAAAACAAAATGGCCAGAAAT 732  
QY 721 CTTTCTCTAAAGTCAAACTGTTGAGTTGATTCAGAGGCCAGTGTAGCGCTTCTAAA 780  
Db 733 CTTTCTCTAAAGTCAAACTGTTGAGTTGATTCAGAGGCCAGTGTAGCGCTTCTAAA 792  
QY 781 CTGAATACCCCAAGGATTTTCAAGATTTGAATTAAGCAAACTAAGAAGAACATGACCAT 840  
Db 793 CTGAATACCCCAAGGATTTTCAAGATTTGAATTAAGCAAACTAAGAAGAACATGACCAT 852  
QY 841 GATGGAAGAACTGACCAATAGTCTGCATATTTATATGGATCTGAGGCCATCAGC 900  
Db 853 GATGGAAGAACTGACCAATAGTCTGCATATTTATATGGATCTGAGGCCATCAGC 912  
QY 901 CAGCTTAAGCAGATGAAGACATCTGCGCAGTCTGTTTGAAGATAACGAAGAACTGAGG 960  
Db 913 CAGCTTAAGCAGATGAAGACATCTGCGCAGTCTGTTTGAAGATAACGAAGAACTGAGG 972  
QY 961 TACTCATTCGGATCTATCAGAGGATGACCATGGTTCGGAATATTTTCATGTCACC 1020  
Db 973 TACTCATTCGGATCTATCAGAGGATGACCATGGTTCGGAATATTTTCATGTCACC 1032  
QY 1021 AACGGCAGATTCCTGCTGGAACCTTGACAATCTCGAGTGACAAATAGTAAACACAC 1080  
Db 1033 AACGGCAGATTCCTGCTGGAACCTTGACAATCTCGAGTGACAAATAGTAAACACAC 1092  
QY 1081 CAGGATGTTTTGGAATTTGAGCCACTTGCCCTACCTTTAGTTACCTGCTATTGAAAGT 1140  
Db 1093 CAGGATGTTTTGGAATTTGAGCCACTTGCCCTACCTTTAGTTACCTGCTATTGAAAGT 1152  
QY 1141 CAGGTTATCGATCGAAGGGCTGTCAGAGAGTTTATTTACCTAAATGATGTCATG 1200  
Db 1153 CAGGTTATCGATCGAAGGGCTGTCAGAGAGTTTATTTACCTAAATGATGTCATG 1212  
QY 1201 TTTGGGAAGGATGCTGGCCAGATGATTTTACACTCACTCCAAAGGCCAGAGTTTAT 1260  
Db 1213 TTTGGGAAGGATGCTGGCCAGATGATTTTACACTCACTCCAAAGGCCAGAGTTTAT 1272  
QY 1261 TTGACATGSCCTGTGCCAAACTGTGCCGAGGGCTGCCAGGTTCTCGGATTAAGGATGGC 1320  
Db 1273 TTGACATGSCCTGTGCCAAACTGTGCCGAGGGCTGCCAGGTTCTCGGATTAAGGATGGC 1332  
QY 1321 TATTGTGACAGGCTTGTAAATAATTCAGCCCTGCGATTCGGATGGTGGGATGCTCTGGA 1380

Db 1333 TATTGTGACAGGCTTGTAAATAATTCAGCCCTGCGATTTGGGATGCTGCTGGA 1392  
QY 1381 AACAGTGGAGGAGTCCGTATATTTCAGGAGGTGAGGTACTGGAGTATTGGAGTTGGA 1440  
Db 1393 AACAGTGGAGGAGTCCGTATATTTCAGGAGGTGAGGTACTGGAGTATTGGAGTTGGA 1452  
QY 1441 CAGCCCTGGCAGTTTGGTGGAGGAATAAACAGATGCTCTTACTGTAATCAGGATGTGCG 1500  
Db 1453 CAGCCCTGGCAGTTTGGTGGAGGAATAAACAGATGCTCTTACTGTAATCAGGATGTGCG 1512  
QY 1501 AATTCTGGCTGCTGTGATAAGTTCTGTGACCAAGCATGCAATGCTTGTCTCTGTGGGTTT 1560  
Db 1513 AATTCTGGCTGCTGTGATAAGTTCTGTGACCAAGCATGCAATGCTTGTCTCTGTGGGTTT 1572  
QY 1561 GATGCTGCGACTCTGGGCAAGATCAATTTTCATGAATTTGATAAAGTGAATCTCTCCCA 1620  
Db 1573 GATGCTGCGACTCTGGGCAAGATCAATTTTCATGAATTTGATAAAGTGAATCTCTCCCA 1632  
QY 1621 AACCAGACTCACTATATTATTCCTCAAAAGGTGAATGCTGCTTATTTTCAGCTTTGCAGAA 1680  
Db 1633 AACCAGACTCACTATATTATTCCTCAAAAGGTGAATGCTGCTTATTTTCAGCTTTGCAGAA 1692  
QY 1681 GTAGCCAAAAGAGGAGTTGAAGTGCCTATAGTGACAAATCCAATPAATTCGACATGCTTCT 1740  
Db 1693 GTAGCCAAAAGAGGAGTTGAAGTGCCTATAGTGACAAATCCAATPAATTCGACATGCTTCT 1752  
QY 1741 ATTGCCAACAGTGGAAACCATCCCTCATATGACAGTGGATGAATGATGCCACCACA 1800  
Db 1753 ATTGCCAACAGTGGAAACCATCCCTCATATGACAGTGGATGAATGATGCCACCACA 1812  
QY 1801 ATACATTTTATCTCAGCTTTCAAAATACAAAGATGAAGAGTTCAAAATGCAGATAACA 1860  
Db 1813 ATACATTTTATCTCAGCTTTCAAAATACAAAGATGAAGAGTTCAAAATGCAGATAACA 1872  
QY 1861 GTGAGGTGGACACAAAGGAGGACCAAACTGAAATTTCTACGCCCAAGAGGTTACGAA 1920  
Db 1873 GTGAGGTGGACACAAAGGAGGACCAAACTGAAATTTCTACGCCCAAGAGGTTACGAA 1932  
QY 1921 AATTTAGTTAGTCCCATTAACACTTCTCCAGAGCGGAAATCCTTTTGAGGATATTTCC 1980  
Db 1933 AATTTAGTTAGTCCCATTAACACTTCTCCAGAGCGGAAATCCTTTTGAGGATATTTCC 1992  
QY 1981 AAAGAAAACGCTTCCCGAAGTTTAAAGACATGATGTTAACTCAACAAGAGAGGCCAG 2040  
Db 1993 AAAGAAAACGCTTCCCGAAGTTTAAAGACATGATGTTAACTCAACAAGAGAGGCCAG 2052  
QY 2041 GAAGAGTGAATTTCCCTGGTGAATATTTTACTCTCTCCAAAAGACGCCCTGAGT 2100  
Db 2053 GAAGAGTGAATTTCCCTGGTGAATATTTTACTCTCTCCAAAAGACGCCCTGAGT 2112  
QY 2101 CTCNAATACCTTGGATTTGCAACTGGACATGGAGACATCACTTTGAAGGATACAATTTG 2160  
Db 2113 CTCNAATACCTTGGATTTGCAACTGGACATGGAGACATCACTTTGAAGGATACAATTTG 2172  
QY 2161 TCCAAGTCAGCCCTGCTGAGATCATTTCTGATGAATCTCACAGCATGCTAAAAATAAAAT 2220  
Db 2173 TCCAAGTCAGCCCTGCTGAGATCATTTCTGATGAATCTCACAGCATGCTAAAAATAAAAT 2232  
QY 2221 CAGCTATAATACAGATGAACAAATGACAGTTTGGTGGCTCCACAGGAAAAACAGTT 2280  
Db 2233 CAGCTATAATACAGATGAACAAATGACAGTTTGGTGGCTCCACAGGAAAAACAGTT 2292  
QY 2281 CATAAAGCATCTGCCAACAGCTTAGGAGTGTCTGAAAGATTCAGAGGTTGACTTTT 2340  
Db 2293 CATAAAGCATCTGCCAACAGCTTAGGAGTGTCTGAAAGATTCAGAGGTTGACTTTT 2352  
QY 2341 CCTGCACTGAGTGTAAAAGTGAATGGTCACTGACCGGGTCAAGATCCACCTGGACTTG 2400  
Db 2353 CCTGCACTGAGTGTAAAAGTGAATGGTCACTGACCGGGTCAAGATCCACCTGGACTTG 2412  
QY 2401 GAGACCACCAAGATTTAGAGTGGAAATCAACCCCAAAACCAATAGCGGAAATG 2460





Db 613 TACTGTGACACAGATAAAGAGTCCCTGGATTAGTGTAATGCAAGATTGGCTTCTCTG 672  
QY 661 AGTGGATTCCACCAACATTCAGGAACAAATCAACTAAACAAATAATGGCCAGAAAT 720  
Db 673 AGTGGATTCCACCAACATTCAGGAACAAATCAACTAAACAAATAATGGCCAGAAAT 732  
QY 721 CTTTCTCTTAAGTCAAACTGTTGCAGTTGTATTTCAGAGGCCAGTGTAGCGCTTCTAAAA 780  
Db 733 CTTTCTCTTAAGTCAAACTGTTGCAGTTGTATTTCAGAGGCCAGTGTAGCGCTTCTAAAA 792  
QY 781 CTGAATTAACCCCAAGGATTTTCAAGATTTGAAGATTAAGCAAACTAAGAGAACATGACCAAT 840  
Db 793 CTGAATTAACCCCAAGGATTTTCAAGATTTGAAGATTAAGCAAACTAAGAGAACATGACCAAT 852  
QY 841 GATGGAAGAACATGACCAATTAAGTTCCTGCATATTATTATFAGGATCTGAGCGCCATCAGC 900  
Db 853 GATGGAAGAACATGACCAATTAAGTTCCTGCATATTATTATFAGGATCTGAGCGCCATCAGC 912  
QY 901 CAGTCTTAAGCAGGATGAAGACATCTCTGCCAGTCTGTTTGAAGATTAAGCAAACTAAGAGAACATGAGG 960  
Db 913 CAGTCTTAAGCAGGATGAAGACATCTCTGCCAGTCTGTTTGAAGATTAAGCAAACTAAGAGAACATGAGG 972  
QY 961 TACTCATTTGGGATCTATFCGAGAGGCATGACCAATGGGTTTCGGAATATTATTCATTTGCACC 1020  
Db 973 TACTCATTTGGGATCTATFCGAGAGGCATGACCAATGGGTTTCGGAATATTATTCATTTGCACC 1032  
QY 1021 AACGGCAGATTCATCTCTGGCTGAACCTTGACAACTCCTCGAGTACAAATAGTAAACAC 1080  
Db 1033 AACGGCAGATTCATCTCTGGCTGAACCTTGACAACTCCTCGAGTACAAATAGTAAACAC 1092  
QY 1081 CAGGATCTTTTTCGAAATTTGACCACTTGCCCTACCTTTAGTTTCACTGCTATTTGAAGT 1140  
Db 1093 CAGGATCTTTTTCGAAATTTGACCACTTGCCCTACCTTTAGTTTCACTGCTATTTGAAGT 1152  
QY 1141 CAGCTTCATCGCATCGAAGGGCTGTCCCAAGAGTTTATTACCTAAATGATGATGTCATG 1200  
Db 1153 CAGCTTCATCGCATCGAAGGGCTGTCCCAAGAGTTTATTACCTAAATGATGATGTCATG 1212  
QY 1201 TTTTGGGAAGGATCTGTGGCCAGATGATTTTACAGTCACTCCAAAGGCCAGAAAGTTTAT 1260  
Db 1213 TTTTGGGAAGGATCTGTGGCCAGATGATTTTACAGTCACTCCAAAGGCCAGAAAGTTTAT 1272  
QY 1261 TTGACATGGCCTGTGCCAAACTGTGCCAGGGCTGCCAGGTTCTCTGGAATTAAGGATGCG 1320  
Db 1273 TTGACATGGCCTGTGCCAAACTGTGCCAGGGCTGCCAGGTTCTCTGGAATTAAGGATGCG 1332  
QY 1321 TATTGTGACAAAGCTTGTATAATTAATTCAGCCTCGGATTTGGGATTTGGGATTTGCTCTGGA 1380  
Db 1333 TATTGTGACAAAGCTTGTATAATTAATTCAGCCTCGGATTTGGGATTTGGGATTTGCTCTGGA 1392  
QY 1381 AACAGTGGAGGAGTCGCTATATTGACAGAGTGGAGGTGAGGTGAGGTGAGGTGAGGTGGA 1440  
Db 1393 AACAGTGGAGGAGTCGCTATATTGACAGAGTGGAGGTGAGGTGAGGTGAGGTGAGGTGGA 1452  
QY 1441 CAGCCCTGCGAGTTTGGTGGAGGAATAAACAGTGTCTTACTGTAAATCAGGATGTCGG 1500  
Db 1453 CAGCCCTGCGAGTTTGGTGGAGGAATAAACAGTGTCTTACTGTAAATCAGGATGTCGG 1512  
QY 1501 AATTCTGCTGCTGATGAAGTTCTGTGACCAAGCATGCAATGCTGTCTGCTGTGGGTTT 1560  
Db 1513 AATTCTGCTGCTGATGAAGTTCTGTGACCAAGCATGCAATGCTGTCTGCTGTGGGTTT 1572  
QY 1561 GATGCTGGGAGTGTGGGCAAGATCATTTTCATGAATTTGATAAGTCACTCTTCTCCCA 1620  
Db 1573 GATGCTGGGAGTGTGGGCAAGATCATTTTCATGAATTTGATAAGTCACTCTTCTCCCA 1632  
QY 1621 AACAGACTCACTATTATTTCAAAAGGTGAATGCTGCTTATTTTTCAGCTTTTGCGAA 1680  
Db 1633 AACAGACTCACTATTATTTCAAAAGGTGAATGCTGCTTATTTTTCAGCTTTTGCGAA 1692  
QY 1681 GTAGCCAAAGAGGAGTTGAAGTGGCTTATAGTGACAATCAATAATTCGACATGCTTCT 1740

Db 1693 GTAGCCAAAGAGGAGTTGAAGGTGCCATAGTGACAATCAATAATTCGACATGCTTCT 1752  
QY 1741 ATTGCCAACAAAGTGGAAACCACTCCACCTCAATPAATGCAGTGGAAATGAATGCCACCACA 1800  
Db 1753 ATTGCCAACAAAGTGGAAACCACTCCACCTCAATPAATGCAGTGGAAATGAATGCCACCACA 1812  
QY 1801 ATACATTTTAAATCTCACGTTTCAAAATCAAAACGATGAAGAGTTTCAAAATCGACATAACA 1860  
Db 1813 ATACATTTTAAATCTCACGTTTCAAAATCAAAACGATGAAGAGTTTCAAAATCGACATAACA 1872  
QY 1861 GTGAGGTGGACACAAAGGAGGACCAAACTGAAATCTACGGCCCAAGGGTTACGAA 1920  
Db 1873 GTGAGGTGGACACAAAGGAGGACCAAACTGAAATCTACGGCCCAAGGGTTACGAA 1932  
QY 1921 AATTTAGTTAGTCCCATAACACTTCTCCAGAGGCGGAAATCCTTTTTCAGGATATTCCC 1980  
Db 1933 AATTTAGTTAGTCCCATAACACTTCTCCAGAGGCGGAAATCCTTTTTCAGGATATTCCC 1992  
QY 1981 AAAGAAAACCGTTCCTCCGAAAGTTTAAAGAGACATGATTTAACTCAACAAGGAGAGCCAG 2040  
Db 1993 AAAGAAAACCGTTCCTCCGAAAGTTTAAAGAGACATGATTTAACTCAACAAGGAGAGCCAG 2052  
QY 2041 GAAAGAGTGAATAATTCCTCTGTAATAATTTTCACTCTCCAAAGAGCCACGTTGAGT 2100  
Db 2053 GAAAGAGTGAATAATTCCTCTGTAATAATTTTCACTCTCCAAAGAGCCACGTTGAGT 2112  
QY 2101 CTCATACCTTTGGATTTGCAACTGGACATGGACATCACTTTTGAAGAGTACAAATTTG 2160  
Db 2113 CTCATACCTTTGGATTTGCAACTGGACATGGACATCACTTTTGAAGAGTACAAATTTG 2172  
QY 2161 TCCAGTCAAGCTTCTGCTGAGATCAATTTCTGATGAACCTCACAGCATGCTAAATAAAAAAT 2220  
Db 2173 TCCAGTCAAGCTTCTGCTGAGATCAATTTCTGATGAACCTCACAGCATGCTAAATAAAAAAT 2232  
QY 2221 CAAGCTATAATAACAGATGAACAAATGACAGTTTGTGTGCTCCACAGGAAACACAGTT 2280  
Db 2233 CAAGCTATAATAACAGATGAACAAATGACAGTTTGTGTGCTCCACAGGAAACACAGTT 2292  
QY 2281 CATAAAAGCATCTTGGCAACAGCTTAGGAGTGTCTGAAAGATTGCAAGAGTTGACATTTT 2340  
Db 2293 CATAAAAGCATCTTGGCAACAGCTTAGGAGTGTCTGAAAGATTGCAAGAGTTGACATTTT 2352  
QY 2341 CCTGAGTGTGTTAAAAGTGAATGTCATGACCAGGTCAGATTCACCCCTGGAGTTG 2400  
Db 2353 CCTGAGTGTGTTAAAAGTGAATGTCATGACCAGGTCAGATTCACCCCTGGAGTTG 2412  
QY 2401 GAGACACAGCAAGATTTAGAGTGGAAACTCACACCCAAACCAATAGCGGAAATGTG 2460  
Db 2413 GAGACACAGCAAGATTTAGAGTGGAAACTCACACCCAAACCAATAGCGGAAATGTG 2472  
QY 2461 ACAAAGAAAGCCCTCATCTCTGATTGTTTCCACTGGAAAGCCAGATGACAAAGAAAG 2520  
Db 2473 ACAAAGAAAGCCCTCATCTCTGATTGTTTCCACTGGAAAGCCAGATGACAAAGAAAG 2532  
QY 2521 AAAATCACAGGAAAGAAAGAGACAGTAGAATGGAGGAAATGCTGAAATACATA 2580  
Db 2533 AAAATCACAGGAAAGAAAGAGACAGTAGAATGGAGGAAATGCTGAAATACATA 2592  
QY 2581 GGGCTTACTGAAGTGTACTTTGGAAGAAAGTGCAGCATTTACACAGATAGTTACTTTGGG 2640  
Db 2593 GGGCTTACTGAAGTGTACTTTGGAAGAAAGTGCAGCATTTACACAGATAGTTACTTTGGG 2652  
QY 2641 TTTTGGCCATGGGAAAGAAAGAAAGTATTTCTCTAGATCTTCTCGAACCAAGAGTCAATG 2700  
Db 2653 TTTTGGCCATGGGAAAGAAAGAAAGTATTTCTAGATCTTCTCGAACCAAGAGTCAATG 2712  
QY 2701 AAGACAAATGGCTTACTTCACTGATACAGAAATAGAGCCAGATACAGAGAGATACA 2760  
Db 2713 AAGACAAATGGCTTACTTCACTGATACAGAAATAGAGCCAGATACAGAGAGATACA 2772  
QY 2761 TTTTGCAGATTTCCCTCAGATATGTAATAAATTTCTAAATAGCAAGTTTGGATTCACATCG 2820  
Db 2773 TTTTGCAGATTTCCCTCAGATATGTAATAAATTTCTAAATAGCAAGTTTGGATTCACATCG 2832

QY 2821 CGGAAAGTCCCTGCTCAGATGCTCAGATGATTCGCGGATTCCTATGCAAGAACTGCAA 2880  
Db 2833 CGGAAAGTCCCTGCTCAGATGCTCAGATGATTCGCGGATTCCTATGCAAGAACTGCAA 2892  
QY 2881 GATATGTTCCCTGAAGAAATTTGACAGAGCTCATTTACAAAGTGGCCATTCTGAGGAT 2940  
Db 2893 GATATGTTCCCTGAAGAAATTTGACAGAGCTCATTTACAAAGTGGCCATTCTGAGGAT 2952  
QY 2941 ATGCAAGTTGGCTTCTCTATATTTTATATCTCATGAGTGCAGTGCGGCACTGGAATATA 3000  
Db 2953 ATGCAAGTTGGCTTCTCTATATTTTATATCTCATGAGTGCAGTGCGGCACTGGAATATA 3012  
QY 3001 TCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTGTCTTGTCTGACAGAGAAATC 3060  
Db 3013 TCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTGTCTTGTCTGACAGAGAAATC 3072  
QY 3061 CGAACACTGGCTACCGAATTCACGAACCTGCCGTTAAAGTTTGCAGGATTTGACAGGCTG 3120  
Db 3073 CGAACACTGGCTACCGAATTCACGAACCTGCCGTTAAAGTTTGCAGGATTTGACAGGCTG 3132  
QY 3121 GAACACATGCTAATAAATTCCTCAAAATGCTTCCTGCTGATATCAGCGAGCTTAATAAT 3180  
Db 3133 GAACACATGCTAATAAATTCCTCAAAATGCTTCCTGCTGATATCAGCGAGCTTAATAAT 3192  
QY 3181 ATTCACCAACTCAGGAATCCTACTATGATCCCAACCTGCCCGTCACTAAAGTCTA 3240  
Db 3193 ATTCACCAACTCAGGAATCCTACTATGATCCCAACCTGCCCGTCACTAAAGTCTA 3252  
QY 3241 GTAACAACTGTAACACAGTAACGTACAAATCCAAAGCATATAAGGACAAAAACAAA 3300  
Db 3253 GTAACAACTGTAACACAGTAACGTACAAATCCAAAGCATATAAGGACAAAAACAAA 3312  
QY 3301 TATAGTTTGAATCATGGAGAGAGAAATCGCTTTTAAATGATTCGTACCAACGTT 3360  
Db 3313 TATAGTTTGAATCATGGAGAGAGAAATCGCTTTTAAATGATTCGTACCAACGTT 3372  
QY 3361 TCTCATGTTGGTGGCAGTTGGATGACATAAGAAAAACCCCTAGGAAGTTTGTGGCTG 3420  
Db 3373 TCTCATGTTGGTGGCAGTTGGATGACATAAGAAAAACCCCTAGGAAGTTTGTGGCTG 3432  
QY 3421 AATGACAACTGACCAACATCAATAAGATGCTCAGACAGTGAAGGCTGTCTCAGGAC 3480  
Db 3433 AATGACAACTGACCAACATCAATAAGATGCTCAGACAGTGAAGGCTGTCTCAGGAC 3492  
QY 3481 TTCTATGAATCCATGTTCCCATACCTTCCCAATTTGAACCTGCCAAGAGATATCGAAAC 3540  
Db 3493 TTCTATGAATCCATGTTCCCATACCTTCCCAATTTGAACCTGCCAAGAGATATCGAAAC 3552  
QY 3541 CGTTTCTCTCATATGATGAGTGCAGGAATGGAGGCTTATCGAGACAAATTTGAAGTAG 3600  
Db 3553 CGTTTCTCTCATATGATGAGTGCAGGAATGGAGGCTTATCGAGACAAATTTGAAGTAG 3612

## RESULT 3

US-09-895-072-20  
; Sequence 20, Application US/09895072  
; Patent No. US2002025550A1

## GENERAL INFORMATION:

; APPLICANT: CANFIELD, WILLIAM M  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
; FILE REFERENCE: 210119USOCNT  
; CURRENT APPLICATION NUMBER: US/09/895,072  
; CURRENT FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; PRIOR APPLICATION NUMBER: US 09/635,872  
; PRIOR FILING DATE: 2000-08-10  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 20  
; LENGTH: 3783  
; TYPE: DNA

## ; ORGANISM: Homo sapiens

US-09-895-072-20

Query Match 96.6%; Score 3476.8; DB 10; Length 3783;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 3484; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 103 AGCCGAGATCAATACCATGTTTGTTCATTCCTATAGACAAATATTCCTGGAAAGTCC 162  
Db 139 AGCCGAGATCAATACCATGTTTGTTCATTCCTATAGACAAATATTCCTGGAAAGTCC 198  
QY 163 TTTCAGAAATCGGCTTTGTCTGCCCATGCCGATTTGACAGTGTGTTTACACCTGGTGAATGGC 222  
Db 199 TTTCAGAAATCGGCTTTGTCTGCCCATGCCGATTTGACAGTGTGTTTACACCTGGTGAATGGC 258  
QY 223 ACAGATCTTGAATCTACTGAAGAACTACAGAGAGTTCAGAGACAGATCGAGAGGAGCAG 282  
Db 259 ACAGATCTTGAATCTACTGAAGAACTACAGAGAGTTCAGAGACAGATCGAGAGGAGCAG 318  
QY 283 AAAGCAATGAGAGAAATCCTTGGGAAACACACGGAACCTACTAAAGAGAGTGAGAG 342  
Db 319 AAAGCAATGAGAGAAATCCTTGGGAAACACACGGAACCTACTAAAGAGAGTGAGAG 378  
QY 343 CAGTTAGAGTGTTCCTAACACACTGCATTAAGGTGCCAATGCTTGTCTTGGACCCAGCC 402  
Db 379 CAGTTAGAGTGTTCCTAACACACTGCATTAAGGTGCCAATGCTTGTCTTGGACCCAGCC 438  
QY 403 CTGCCAGCAACATCACCCTGAAGGACCTGCCATCTCTTTATCTCTTTTCATTCTGCC 462  
Db 439 CTGCCAGCAACATCACCCTGAAGGACCTGCCATCTCTTTATCTCTTTTCATTCTGCC 498  
QY 463 AGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTACCACATGCTCAGTTGTTGTT 522  
Db 499 AGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTACCACATGCTCAGTTGTTGTT 558  
QY 523 TTTGACAGTACTAAGGATGTTGAAGATGCCACTCTGGACTGCTTTAAAGGAAATAGCAGA 582  
Db 559 TTTGACAGTACTAAGGATGTTGAAGATGCCACTCTGGACTGCTTTAAAGGAAATAGCAGA 618  
QY 583 CAGACAGTATGAGGGGCTACTTGACACAGATAAAGAGTCCCTGGATTTAGTGTGAATG 642  
Db 619 CAGACAGTATGAGGGGCTACTTGACACAGATAAAGAGTCCCTGGATTTAGTGTGAATG 678  
QY 643 CAAGATTTGGCTTTCTCTGAGTGGATTTCCACCAACATTCAGAGAAACAAATCACTAAA 702  
Db 679 CAAGATTTGGCTTTCTCTGAGTGGATTTCCACCAACATTCAGAGAAACAAATCACTAAA 738  
QY 703 ACAAAATGCCAGAAATCTTCT 762  
Db 739 ACAAAATGCCAGAAATCTTCT 798  
QY 763 AGTGAGCCCTTCTAAACATGAATTAACCCCAAGGATTTTCAAGATTTGAATTAAGCAACT 822  
Db 799 AGTGAGCCCTTCTAAACATGAATTAACCCCAAGGATTTTCAAGATTTGAATTAAGCAACT 858  
QY 823 AAGAAGAACTGACCATTTGATGAAAGAACTGACATAAGTCCCTGCATATTTATTTATGG 882  
Db 859 AAGAAGAACTGACCATTTGATGAAAGAACTGACATAAGTCCCTGCATATTTATTTATGG 918  
QY 883 GATCTGAGCGCATCAGCCAGTCTTAAGCAGGATGAAGACATCTCTGCCAGTCTTTTGA 942  
Db 919 GATCTGAGCGCATCAGCCAGTCTTAAGCAGGATGAAGACATCTCTGCCAGTCTTTTGA 978  
QY 943 GATAACGAAACTGAGGTACTCTATTCGATCTATCGAGAGCATGCACCATGGTTCGG 1002  
Db 979 GATAACGAAACTGAGGTACTCTATTCGATCTATTCGAGAGCATGCACCATGGTTCGG 1038  
QY 1003 AATATTTTCTATGTCACCAACGGGAGATTTCCATCTCTGGCTGAACCTTGACAACTCCTGA 1062  
Db 1039 AATATTTTCTATGTCACCAACGGGAGATTTCCATCTCTGGCTGAACCTTGACAACTCCTGA 1098  
QY 1063 GTGACAAATAGTAACACACAGGATGTTTTCGAAATTTTGAGCACTTGCCTACCTTAGT 1122

Db 1099 GTGCAATAGTAACACACCAGGATGTTTTTGGAAATTTGAGCCACTTGCCTACCTTAGT 1158  
QY 1123 TCACCTGCTATTGAAAGTCACGTTTCATCGCATCGAAGGCTGTCCCAAGAGTTTATTAC 1182  
Db 1159 TCACCTGCTATTGAAAGTCACGTTTCATCGCATCGAAGGCTGTCCCAAGAGTTTATTAC 1218  
QY 1183 CTAATGATGATGTCATGTTTGGGAAGGATGCTCGGCAGATGATTTTACAGTCACTCC 1242  
Db 1219 CTAATGATGATGTCATGTTTGGGAAGGATGCTCGGCAGATGATTTTACAGTCACTCC 1278  
QY 1243 AAAGCCAGAGGTTTATTGACATGGCCCTGTGCCAAACTGTGCCGAGGCTGCCCAGGT 1302  
Db 1279 AAAGCCAGAGGTTTATTGACATGGCCCTGTGCCAAACTGTGCCGAGGCTGCCCAGGT 1338  
QY 1303 TCCTGGATTAAAGATGCTATTGTGACAAAGGCTTGTAAATAATTCACGCTGCCATTTGGGAT 1362  
Db 1339 TCCTGGATTAAAGATGCTATTGTGACAAAGGCTTGTAAATAATTCACGCTGCCATTTGGGAT 1398  
QY 1363 GTTGGGATTTGCTCTGGAACAGTGGAGGAGTGCCTATATTGACGAGGTTGGAGTACT 1422  
Db 1399 GTTGGGATTTGCTCTGGAACAGTGGAGGAGTGCCTATATTGACGAGGTTGGAGTACT 1458  
QY 1423 GGGAGTATTGAGCTTGGACAGCCCTGGCAGTTTGGTGGAGGAATAAACAGTGTCTCTTAC 1482  
Db 1459 GGGAGTATTGAGCTTGGACAGCCCTGGCAGTTTGGTGGAGGAATAAACAGTGTCTCTTAC 1518  
QY 1483 TGTAAATCAGGGATGTGCGAATTCCTGGCTCGCTGATAAGTCTGTGACCAAGCATGCAAT 1542  
Db 1519 TGTAAATCAGGGATGTGCGAATTCCTGGCTCGCTGATAAGTCTGTGACCAAGCATGCAAT 1578  
QY 1543 GTCTTGTCTGTGGGTTTGATGCTGGCGACTGTGGCAAGATCAATTTTCATGAATTTGAT 1602  
Db 1579 GTCTTGTCTGTGGGTTTGATGCTGGCGACTGTGGCAAGATCAATTTTCATGAATTTGAT 1638  
QY 1603 AAAGTATCTCTTCCCAACACAGACTCACTATATATTATTCAAAAGTGTAATGCTGCT 1662  
Db 1639 AAAGTATCTCTTCCCAACACAGACTCACTATATATTATTCAAAAGTGTAATGCTGCT 1698  
QY 1663 TATTTACGCTTTGCAGAGTAGCCAAAAGAGGAGTTGAAGTGCTCTATAGTGACAATCCA 1722  
Db 1699 TATTTACGCTTTGCAGAGTAGCCAAAAGAGGAGTTGAAGTGCTCTATAGTGACAATCCA 1758  
QY 1723 ATAATTCGACATGCTTCTATTGCCAAACAGTGGAAAACCATCCACCTCAATAGTCAAGT 1782  
Db 1759 ATAATTCGACATGCTTCTATTGCCAAACAGTGGAAAACCATCCACCTCAATAGTCAAGT 1818  
QY 1783 GGAATGAATGCCACCACAAATACATTTTAACTCTCAGTTTCAAATACAAACGATGAAG 1842  
Db 1819 GGAATGAATGCCACCACAAATACATTTTAACTCTCAGTTTCAAATACAAACGATGAAG 1878  
QY 1843 TTCAAATGCGAGATAACAGTGGAGTGGACACAAAGGGAGGACCAAACTGAATTTCTACG 1902  
Db 1879 TTCAAATGCGAGATAACAGTGGAGTGGACACAAAGGGAGGACCAAACTGAATTTCTACG 1938  
QY 1903 GCCAGAGAGGTTACGAAAATTTAGTTAGTCCCATACACTTCTCCAGAGGCGGAAATC 1962  
Db 1939 GCCAGAGAGGTTACGAAAATTTAGTTAGTCCCATACACTTCTCCAGAGGCGGAAATC 1998  
QY 1963 CTTTTTGGAGTATTTCCCAAGAAAACGCTTCCCGAAGTTTAAAGACACATGATTTAAC 2022  
Db 1999 CTTTTTGGAGTATTTCCCAAGAAAACGCTTCCCGAAGTTTAAAGACACATGATTTAAC 2058  
QY 2023 TCACAGGAGGAGGCCAGGAGGTTGAAAATTTCCCTGGTAAATATTTTCACCTCTTCCA 2082  
Db 2059 TCACAGGAGGAGGCCAGGAGGTTGAAAATTTCCCTGGTAAATATTTTCACCTCTTCCA 2118  
QY 2083 AAAGACCCAGTTGAGTCTCAATACCTTGGATTTGCACTGGAACATGGAGACATCACT 2142  
Db 2119 AAAGACCCAGTTGAGTCTCAATACCTTGGATTTGCACTGGAACATGGAGACATCACT 2178  
QY 2143 TTGAAGGATACAAATTTGTCCAACTAGCCTTGTGATGATCAATTTCTGATGAACACTCAC 2202  
Db 2179 TTGAAGGATACAAATTTGTCCAACTAGCCTTGTGATGATCAATTTCTGATGAACACTCAC 2238

QY 2203 CATGCTAAATAAAAATCAAGCTATAATACAGATGAAACAAATGACAGTTTGGTGCT 2262  
Db 2239 CATGCTAAATAAAAATCAAGCTATAATACAGATGAAACAAATGACAGTTTGGTGCT 2298  
QY 2263 CCACAGGAAACACAGGTTTCATAAAGAGCATCTTGGCCAAACAGCTTAGGAGTGTCTGAAAGA 2322  
Db 2299 CCACAGGAAACACAGGTTTCATAAAGAGCATCTTGGCCAAACAGCTTAGGAGTGTCTGAAAGA 2358  
QY 2323 TTGCAGAGTTTGACATTTTCCCTGAGTGTGTAAGGTGAATGTCATGACACAGGTCAG 2382  
Db 2359 TTGCAGAGTTTGACATTTTCCCTGAGTGTGTAAGGTGAATGTCATGACACAGGTCAG 2418  
QY 2383 AATCCACCCCTGGACTTTGGAGACACACAGATTTAGAGTGGAAACTCACACCCCAAAA 2442  
Db 2419 AATCCACCCCTGGACTTTGGAGACACACAGATTTAGAGTGGAAACTCACACCCCAAAA 2478  
QY 2443 ACCATAGCGGAAATGTGACAAAAGAAAGCCCATCTCTGATTTGTTCACCTGGAAGC 2502  
Db 2479 ACCATAGCGGAAATGTGACAAAAGAAAGCCCATCTCTGATTTGTTCACCTGGAAGC 2538  
QY 2503 CAGATGACAAAAGAAAGAAATCACAGGGAAGAAAGAGACAGTACAGTGGAGGAA 2562  
Db 2539 CAGATGACAAAAGAAAGAAATCACAGGGAAGAAAGAGACAGTACAGTGGAGGAA 2598  
QY 2563 AATGCTGAAAATACATAGGCGTTTACTGAAAGTGTACTTTGGAAGAAAGCTGCAGCATTTAC 2622  
Db 2599 AATGCTGAAAATACATAGGCGTTTACTGAAAGTGTACTTTGGAAGAAAGCTGCAGCATTTAC 2658  
QY 2623 ACAGATAGTTACTTTGGCGCTTTTGGCAATGGGAGAAAGAAAGTATTTCCCTAGATCTCTC 2682  
Db 2659 ACAGATAGTTACTTTGGCGCTTTTGGCAATGGGAGAAAGAAAGTATTTCCCTAGATCTCTC 2718  
QY 2683 GACGAAGAGAGTCAATTTGAAGACACATTTGGCCTACTTTCATCATAGCAGATAGAGCC 2742  
Db 2719 GACGAAGAGAGTCAATTTGAAGACACATTTGGCCTACTTTCATCATAGCAGATAGAGCC 2778  
QY 2743 AGATACAGAGAGATACATTTGCGAGATTTCCCTCAGATATGTAATAATAAATTTCTAAATAGC 2802  
Db 2779 AGGCACTAAAGATACATTTGCGAGATTTCCCTCAGATATGTAATAATAAATTTCTAAATAGC 2838  
QY 2803 AAGTTTGGATTACATCGCGGAAAGTCCCTGCTCAGATGCTGCTCAGATGATGACCGGATTT 2862  
Db 2839 AAGTTTGGATTACATCGCGGAAAGTCCCTGCTCAGATGCTGCTCAGATGATGACCGGATTT 2898  
QY 2863 GTTATGCAAGAACTGCAAGATATGTTCCCTGAAAGATTTTGAAGAACTGCAAGCTCAATTTCAAAA 2922  
Db 2899 GTTATGCAAGAACTGCAAGATATGTTCCCTGAAAGATTTTGAAGAACTGCAAGCTCAATTTCAAAA 2958  
QY 2923 GTGCGCATTTGAGGATGATGAGTTTGGCTTCTCTTATTTTATCTCATGAGTGA 2982  
Db 2959 GTGCGCATTTGAGGATGATGAGTTTGGCTTCTCTTATTTTATCTCATGAGTGA 3018  
QY 2983 GTGCGCCACTGAATATATCTCAAGTCTTTGATGAGTTGATACAGATCAATCTGGTGTG 3042  
Db 3019 GTGCGCCACTGAATATATCTCAAGTCTTTGATGAGTTGATACAGATCAATCTGGTGTG 3078  
QY 3043 TTGCTGACAGAAATCCGAACACTGGCTACGAAATTCAGAACTGCGGTTAAGTTTG 3102  
Db 3079 TTGCTGACAGAAATCCGAACACTGGCTACGAAATTCAGAACTGCGGTTAAGTTTG 3138  
QY 3103 CAGGATTTGACAGGTTCTGGAACACATGCTTAATAAATGCTCAAAAATGCTTCTGCTGAT 3162  
Db 3139 CAGGATTTGACAGGTTCTGGAACACATGCTTAATAAATGCTCAAAAATGCTTCTGCTGAT 3198  
QY 3163 ATCAGCAGCTAAATAATATCCCAACTCAGGAATCCTACTATGATCCCAACTGGCA 3222  
Db 3199 ATCAGCAGCTAAATAATATCCCAACTCAGGAATCCTACTATGATCCCAACTGGCA 3258  
QY 3223 CCGGTCACTAAAGTCTAGTAACAACTGTAACAGTAAACAGTAACTGACAAATCCCAAGCA 3282  
Db 3259 CCGGTCACTAAAGTCTAGTAACAACTGTAACAGTAAACAGTAACTGACAAATCCCAAGCA 3318









## GENERAL INFORMATION:

; APPLICANT: CANFIELD, WILLIAM M  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
; FILE REFERENCE: 21019US00NT  
; CURRENT APPLICATION NUMBER: US/09/895,072  
; CURRENT FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; PRIOR APPLICATION NUMBER: US 09/635,872  
; PRIOR FILING DATE: 2000-08-10  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 5597  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-895-072-4

Query Match 96.3%; Score 3465.6; DB 10; Length 5597;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 3477; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 103 ACCGGAGATCAATACCATGTTTGGTTGATTCCTATAGAGACAATATTCCTGGAAGTCC 162  
DB 297 AGCCGAGATCAATACCATGTTTGGTTGATTCCTATAGAGACAATATTCCTGGAAGTCC 356  
QY 163 TTTTCAGAAATCGGCTTTGCTGCCATGCCGATGAGGTTGTTTACACCTGGGTGAATGCC 222  
DB 357 TTTTCAGAAATCGGCTTTGCTGCCATGCCGATGAGGTTGTTTACACCTGGGTGAATGCC 416  
QY 223 ACAGATCTTGAATCTAGTGAAGAACTACAGAGGTGAGAGACAGATGAGGAGGAGCAG 282  
DB 417 ACAGATCTTGAATCTAGTGAAGAACTACAGAGGTGAGAGACAGATGAGGAGGAGCAG 476  
QY 283 AAGCAATGAGAGAAATCTTGGGAAACACACAGCAACGAACTACTAAGAGAGTGAAG 342  
DB 477 AAGCAATGAGAGAAATCTTGGGAAACACACAGCAACGAACTACTAAGAGAGTGAAG 536  
QY 343 CAGTTAGATGTTGCTAACACACTGCAATTAAGGTGCCAATGCTTGTCTGAGACCCAGCC 402  
DB 537 CAGTTAGATGTTGCTAACACACTGCAATTAAGGTGCCAATGCTTGTCTGAGACCCAGCC 596  
QY 403 CTGCCAGCCAACTACCTGAGGACCTGCAATCTCTTATPCCTTTCTTCAATCTGCC 462  
DB 597 CTGCCAGCCAACTACCTGAGGACCTGCAATCTCTTATPCCTTTCTTCAATCTGCC 656  
QY 463 AGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGCTCAGTTGTTGT 522  
DB 657 AGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGCTCAGTTGTTGT 716  
QY 523 TTTGACAGTACTAAGATGTTGAAGATGCCACTCTGGACTGCTTAAAGGAAATAGCAGA 582  
DB 717 TTTGACAGTACTAAGATGTTGAAGATGCCACTCTGGACTGCTTAAAGGAAATAGCAGA 776  
QY 583 CAGACAGTATGAGGGCTACTTGACACAGATAAAGAGTCCCTGGATTAAGTCTAATG 642  
DB 777 CAGACAGTATGAGGGGGTACTTGACACAGATAAAGAGTCCCTGGATTAAGTCTAATG 836  
QY 643 CAAGATTTGGCTTTCTTCTGAGTGGATTTCCACCAATTCAGGAAACAAATCAACTAAA 702  
DB 837 CAAGATTTGGCTTTCTTCTGAGTGGATTTCCACCAATTCAGGAAACAAATCAACTAAA 896  
QY 703 ACAAAATGCGCAAAATCTTTCTCTTAAAGTCAAACTGTTGCAAGTGTATTCAGAGGCC 762  
DB 897 ACAAAATGCGCAAAATCTTTCTCTTAAAGTCAAACTGTTGCAAGTGTATTCAGAGGCC 956  
QY 763 AGTGTAGCGCTTCTAAACTGAATAACCCCAAGGATTTTCAAGATTTGAATTAACCAACT 822  
DB 957 AGTGTAGCGCTTCTAAACTGAATAACCCCAAGGATTTTCAAGATTTGAATTAACCAACT 1016  
QY 823 AAGAGAAATGACCATTTGATGGAAGAACTGACCATTAAGTCTCTGCAATTTTATTTATGG 882  
DB 1017 AAGAGAAATGACCATTTGATGGAAGAACTGACCATTAAGTCTCTGCAATTTTATTTATGG 1076

QY 883 GATCTGAGCGCATCAGCAGCTCTTAAGCAGGATGAAGACATCTCTGCGCAGTCTGTTTGA 942  
DB 1077 GATCTGAGCGCATCAGCAGCTCTTAAGCAGGATGAAGACATCTCTGCGCAGTCTGTTTGA 1136  
QY 943 GATACGAGAACTGAGTACTCATTTGGATCTATPCGAGAGGATGCAACCATGGTTCGG 1002  
DB 1137 GATAACGAGAACTGAGTACTCATTTGGATCTATPCGAGAGGATGCAACCATGGTTCGG 1196  
QY 1003 AATATTTTCAATGTCACCAAGGCGAGATTCATCTCGCTGGCTGAACCTTGACAAATCCTCGA 1062  
DB 1197 AATATTTTCAATGTCACCAAGGCGAGATTCATCTCGCTGGCTGAACCTTGACAAATCCTCGA 1256  
QY 1063 GTGCAATAGTAACACACAGGATGTTTTCGAAATTTGAGCCACTTGCCTACCTTAGT 1122  
DB 1257 GTGCAATAGTAACACACAGGATGTTTTCGAAATTTGAGCCACTTGCCTACCTTAGT 1316  
QY 1123 TCACCTGCTATTTGAAGTCAGTTCATCGCATCGAAGGCTGTCCAGAAAGTTTATTATAC 1182  
DB 1317 TCACCTGCTATTTGAAGTCAGTTCATCGCATCGAAGGCTGTCCAGAAAGTTTATTATAC 1376  
QY 1183 CTAATGATGATGTCATGTTTGGGAAGGATGCTGCGCCAGATGATTTTACAGTCACTCC 1242  
DB 1377 CTAATGATGATGTCATGTTTGGGAAGGATGCTGCGCCAGATGATTTTACAGTCACTCC 1436  
QY 1243 AAAGCCAGAAAGTTTATTTGACATGCGCTGTGCCAAACTGTGCCGAGGGCTGCCAGGT 1302  
DB 1437 AAAGCCAGAAAGTTTATTTGACATGCGCTGTGCCAAACTGTGCCGAGGGCTGCCAGGT 1496  
QY 1303 TCCTGGATTAAGGATGCTATTTGTGACAAGGCTTGTATTAATTTACAGCTGGATGGAT 1362  
DB 1497 TCCTGGATTAAGGATGCTATTTGTGACAAGGCTTGTATTAATTTACAGCTGGATGGAT 1556  
QY 1363 GGTGGGATTTGCTCTGGAACACAGTGGAGGAGTGCCTATATTTGAGGAGGTGGAGTACT 1422  
DB 1557 GGTGGGATTTGCTCTGGAACACAGTGGAGGAGTGCCTATATTTGAGGAGGTGGAGTACT 1616  
QY 1423 GGGAGTATTTGGATTTGACAGCCCTGCGAGTTTGGTGGAGAAATAAACAGTCTCTTAC 1482  
DB 1617 GGGAGTATTTGGATTTGACAGCCCTGCGAGTTTGGTGGAGAAATAAACAGTCTCTTAC 1676  
QY 1483 TGTAAATCAGGATGTCGGAATTTCTCGCTCGCTGATGAAGTTCTGTGACCAAGCATGCAAT 1542  
DB 1677 TGTAAATCAGGATGTCGGAATTTCTCGCTCGCTGATGAAGTTCTGTGACCAAGCATGCAAT 1736  
QY 1543 GTCTTTGCTGTGGTTTGTGATGCTGGGCAAGTGTGGGCAAGTCAATTTTCAATTTGAT 1602  
DB 1737 GTCTTTGCTGTGGTTTGTGATGCTGGGCAAGTGTGGGCAAGTCAATTTTCAATTTGAT 1796  
QY 1603 AAAGTGTATCTTCTCCCAACACAGACTCATATATTTCCAAAGGTGAATGCCTGCCT 1662  
DB 1797 AAAGTGTATCTTCTCCCAACACAGACTCATATATTTCCAAAGGTGAATGCCTGCCT 1856  
QY 1663 TATTTACAGCTTTGACAGAACTAGCCAAAGAGGAGTTGAAGTGCCTTATAGTGACAAATCCA 1722  
DB 1857 TATTTACAGCTTTGACAGAACTAGCCAAAGAGGAGTTGAAGTGCCTTATAGTGACAAATCCA 1916  
QY 1723 ATAAATTCGACATGCTTCTATTTGCGCAACAGTGGGAAACCACTCATATATGACAGT 1782  
DB 1917 ATAAATTCGACATGCTTCTATTTGCGCAACAGTGGGAAACCACTCATATATGACAGT 1976  
QY 1783 GGAATGAATGCCACCAATACATTTTAACTCACCCTTTCAAAATACAAACAGTGAAGAG 1842  
DB 1977 GGAATGAATGCCACCAATACATTTTAACTCACCCTTTCAAAATACAAACAGTGAAGAG 2036  
QY 1843 TTCAAAATGCAGATAACAGTGGAGGTGGACACAAAGGAGGACCAAACTGAATTTCTAGC 1902  
DB 2037 TTCAAAATGCAGATAACAGTGGAGGTGGACACAAAGGAGGACCAAACTGAATTTCTAGC 2096  
QY 1903 GCCCAGAAAGGTTACCAAAATTTAGTTAGTCCCATTAACACTTCTTCCAGAGGCGGAAATC 1962  
DB 2097 GCCCAGAAAGGTTACCAAAATTTAGTTAGTCCCATTAACACTTCTTCCAGAGGCGGAAATC 2156











Db 2785 AGCAAGTTGGATTACATCCAGAAAGTCCCTGCACACATGCCGCACATGATTGACAGG 2844  
QY 2860 ATTGTTATGCAAGAACTGCAAGATATGTTCCCTGAAGAAATTTGACAAGACGTCATTTTAC 2919  
Db 2845 ATCGTTATGCAAGAACTGCAAGATATGTTCCCTGAAGAAATTTGACAAGACGTCATTTTAC 2904  
QY 2920 AAAGTGGCCCATCTCAGGATATGCAAGTTGGCTTCTCTATTTATTTATTTATCTCATGAGT 2979  
Db 2905 AAGTGGCTGACTCAGACATGCAAGTTGGCTTCTCTATTTATTTATTTATCTCATGAGT 2964  
QY 2980 GCAGTGCAGCCATGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGT 3039  
Db 2965 GCAGTTGACGCCCTCAATATTTCCCAAGTCTTTTCATGAAGTACAGACACAGCAATCTGGT 3024  
QY 3040 GTCTTTGTCAGACAGAAATCCGACACATGGCTACCAGAAATTCAGGAATGCGGTTAAGT 3099  
Db 3025 GTCTTTGTCATAGGAAATCCGACACATGGCCACGAGAAATTCAGGACCTACCTTTAAGC 3084  
QY 3100 TTGACGATTTGACAGGTTGGAACACATGCTAAATAATTTGCTCAAAAATGCTTCTGCT 3159  
Db 3085 TTGACGATTTGACAGGTTGGAACACATGCTAAATAATTTGCTCAAAAATGCTTCTGCT 3144  
QY 3160 GATATCACGACGATTAATATATTCACCAACTCAGGAATCTCTATGATCCCAACCTG 3219  
Db 3145 AATATCACTCAACTCAACACATCCACCGACTCAGGAAGCATACTACGACCCCAACCTG 3204  
QY 3220 CCACGGTCTCACTAAAGTCTAGTAACAACTGTAACCACTGTAACCACTGTAACCACTGTAACCA 3279  
Db 3205 CTTCCGGTCACTAAGAGTCTGTGCACTGTAAGCCAGTGAAGCCAGTGAAGCCAGTGAAGCCAGT 3264  
QY 3280 GCATATAAGGCAAAAACAAATATAGTTTGAATCATGGGAGAGAAATTCGCTTTT 3339  
Db 3265 GCCTATAAGACAAAGCAAAATACAGTTTGAATCATGGGAGAGAAATTCGCTTTT 3324  
QY 3340 AAAATGATTCGACCAAGCTTCTCATGTTGGTGGCCAGTGGAGACATAGAAAAC 3399  
Db 3325 AAGATATAGCAACCAATGTTTCTCATGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 3384  
QY 3400 CTTAGGAAGTTGTTGCTGCTGAATGACAACTGACCAATGACCAATGACCAATGACCAATGACCA 3459  
Db 3385 CCAGGAGAGTTCTGTTGCTGAATGACAACTGACCAATGACCAATGACCAATGACCAATGACCA 3444  
QY 3460 GTGAAGCTTTCTCAGGAGTCTATGATCCATGATCCCATACCTTTCCCAATTTGAA 3519  
Db 3445 GTGAAGCTTTCTCAGGAGTCTATGATCCATGATCCCATACCTTTCCCAATTTGAG 3504  
QY 3520 CTGCAAGAGATGATCAAGCTTTTCTCATATGATGATGATGATGATGATGATGATGATGATGAT 3579  
Db 3505 CTGCAAGAGATGATCAAGCTTTTCTCATATGATGATGATGATGATGATGATGATGATGATGAT 3564  
QY 3580 TATCGAGACAAATTTGAAT 3598  
Db 3565 TATCGAGACAACTGAAT 3583

## RESULT 8

US-09-986-552-16  
; Sequence 16, Application US/09986552  
; Patent No. US20020150981A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
; FILE REFERENCE: 215089US77DIV  
; CURRENT APPLICATION NUMBER: US/09/986,552  
; CURRENT FILING DATE: 2001-11-09  
; PRIOR APPLICATION NUMBER: 09/635,872  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 16  
; LENGTH: 5229

; TYPE: DNA  
; ORGANISM: Mus musculus  
US-09-986-552-16

Query Match 65.6%; Score 2362.6; DB 10; Length 5229;  
Best Local Similarity 81.2%; Pred. No. 0;  
Matches 2842; Conservative 1; Mismatches 590; Indels 56; Gaps 6;

QY 103 AGCCGAGATCAATACCATGTTTCTGTTGATTCTCTATAGACACATATGCTGGAAGTCC 162  
Db 148 AGCCGAGATCAATACCATGTTTCTGTTGATTCTCTATAGACACATATGCTGGAAGTCC 207  
QY 163 TTTTCAAGATCGGCTTGTCTGCCCATGCCGATTGACCTTGTACACCTGGGTGAATGCG 222  
Db 208 TTTTCAAGATCGGCTTGTCTGCCCATGCCGATTGACCTTGTACACCTGGGTGAATGCG 267  
QY 223 ACAGATCTTGAACCTGCTGAAGGAACTACAGAGGTGACAGACAGATGGAGGAGGACG 282  
Db 268 ACTGACCTTGAACCTGCTGAAGGAACTACAGAGGTGACAGAGGTGACAGGAGGAGGACG 327  
QY 283 AAAGCAATGAGAGAAATCCCTTGGGAAAACACACACGAACTTACTAAGAAAGTGAAG 342  
Db 328 AGAGCCATGCGGGAACCCCTCGGGAAGAACACACACGAACTTACTAAGAAAGTGAAG 387  
QY 343 CAGTTAGAGTGTGCTTAACACACTGCAATTAAGGTGCAATGCTTGTCTTGTTCATCTGCC 402  
Db 388 CAGCTGGAATGCTGCTGACGCACTGCAATTAAGGTGCAATGCTTGTCTTGTTCATCTGCC 447  
QY 403 CTGCCAGCAACACTACCTGGAAGGACCTGCGCATCTCTTATCTCTTCTTCTTCTGCC 462  
Db 448 CTGCCAGCAACACTACCTGGAAGGATCTGCAACACCTTACCCTATCTTCCAGCGTCC 507  
QY 463 AGTGACATTTTCAATGTTGCAAAACCAAAACCTTCTACCAATGCTCAGTTGTGTT 522  
Db 508 AGGACATGTTCAATGTTGCAAAACCAAAACCTTCTACCAATGCTCAGTTGTGTT 567  
QY 523 TTTGACAGTACTAAGAGTGTGGAAGTGGCCACTCTGCACTGCTTAAAGGAATAGCAGA 582  
Db 568 TTTGACACTACTAAGAGTGTGGAAGGCGCCATCTGCTGAGCGCTTAAAGGAGGCGCAGCA 627  
QY 593 CAGACAGTATGAGGGGTACTTGCACACAGATAAAGTCCCTGATGATGCTAATG 642  
Db 628 ACAGATGTTTGAGAGGCTACTTGCACACAGATAAAGTCCCTGATGATGCTAATG 687  
QY 643 CAAGATTTGGCTTTCTCAGTGGATTTCCACCAACATTTCAAGGAACAAATCAACTAAAA 702  
Db 688 CAAGCTTGGCGTTCTCAGTGGATTTCCACCGACCTTCAAGGAGAGAGTCAACTGAAG 747  
QY 703 ACAGATTTGCCAGAAATCTTCTCTCT---AAAGTCAACTGTTGCAAGTGTGATTCAGAG 759  
Db 748 ACAAGCTGCCAAGAAAGCTTTCCCTCTAAAAATAAAGCTTTCGCGCTGATCTCGGAG 807  
QY 760 GCCAGTGTAGGCTTCTAAAACTCAATAACCCCAAGGATTTTCAAGAAATGAATAAGCAA 819  
Db 808 GCCAGTGTAGGCTTCTTGAATTTGAATTTCCCAAGGTTTCCCAAGAGTGAACACAGCAG 867  
QY 820 ACTAAGAGAAACATGACCATTTGATGGAAAAGAACTGACCAATAGTCTCGCATATTTA 879  
Db 868 ACCAAGAGAAACATGACCATTTGATGGAAAAGAACTGACCAATAGTCTCGCATATTTA 927  
QY 880 TGGATCTGAGCGGCATCAGCCATCTAAGCAGGATGAACACATCTCTGCCAGTCTGTTT 939  
Db 928 TGGACCTTGAAGTGGCATCAGCCATCTCAAGCAGGATGAGGACGCTTGGCAGCGGCTT 987  
QY 940 GAAGATAACGAAGAACTGAGTACTCATTTGCGATCTATCGAGAGGCAATGCCATGGGTT 999  
Db 988 GAGGATATGAGAGCTGAGGTACTCGCTGGATCTATCGAGAGACAGCGCCATGGGA 1047  
QY 1000 CGGAATATTTTCAATGTCACCAAGGCGGAGATTCCTGCTGCTGAACCTTTGACAACTCT 1059  
Db 1048 CGGAATATTTTCAATGTCACCAAGGCGGAGATTCCTGCTGCTGCTGAACCTTTGACAACTCT 1107  
QY 1060 CGAGTGACAATAGTAACACACACAGGATGCTTTTTCGAAATTTGAGCCACTTGCCTACCTT 1119



Db	1108	CGAGTGACCAATGAGTGCACCCACGAGGACATTTTCCAAAATCTGAGCCACTGGCTACTTTC	1167
QY	1120	AGTTTCACTGCTATTGAAAGTCACGTTTCATCGCATCGGAAGGCTCTCCCAAGCTTTATT	1179
Db	1168	AGTTCCCTGCTATTGAAAGTCACATTCACCGATCGAAGGCTCTCCCAAGCTTTATT	1227
QY	1180	TACTAAATGATGATGTCATGTTTGGGAAGGATGTCTGGCCAGATGATTTTTACAGTCAC	1239
Db	1228	TATCTAAATGACGATGTCATGTTTCGGTAAGGACGCTCTGCGCGGACGATTTTTACAGCCAC	1287
QY	1240	TCCAAAGCCACGAAGGTTTATTTTGACATGCGCTGTGCCAAACTGTGCGGAGGCTGCGCCA	1299
Db	1288	TCCAAAGCTCAAAGGTTTATTTTGACATGCGCTGTGCCAAACTGTGCGAGAGGCTGCGCGC	1347
QY	1300	GGTTCCTGGATTAAAGGATGGCTATTGTGCACAAAGGCTTGTAATAATTCAAGCTCGGATTGG	1359
Db	1348	GGCTCCTGGATAAAGGACGCTATTGTGATAGGCGCTGTATACCTCACCTGTGACTGG	1407
QY	1360	GATGGTGGGATTGCTCTGGAACAGTGAGGAGTGCTGTATATTGCAGGAGGTGGAGGT	1419
Db	1408	GATGGCGGAAACTGCTCTGTGTAATCTGCAGGGAACCGGTTGTGCAAGAGGTGGGSGT	1467
QY	1420	ACTGGGAGTATCGAGTGTGCAGAGCCCTGGCAGTTTGGTGAGAGTAATAACAGTGTCTCT	1479
Db	1468	ACCGGGAATATTGGAGCTGGACAGCACTGGCAGTTTGGTGAGGAAATAACACCATCTCT	1527
QY	1480	TACTGTAAATCAGGAGTGTGCGAATTCTTGGCTGCGCTGATAAGTTCTGTGCACCAAGCATGC	1539
Db	1528	TACTGTAAACCAAGATGTGCAAACTCTGGCTGGCTGACAAGTTCTGTGCACCAAGCCCTGT	1587
QY	1540	AATGTCTTGTCTGTGGGTTTGATGCTGGCAGCTGTGGCAAGATCATTTTCATGAATTG	1599
Db	1588	AACGCTTTATPCTTCGCGGTTTGATGCTGGTACTGTGCACAAGATCATTTTCATGAATTG	1647
QY	1600	TATAAAGTATCTCTCCCAACACAGACTCACTATATTATTCCAAAGGTGAATGCCTG	1659
Db	1648	TATAAAGTAACTCTCTCCCAACACAGACTCACTATGTTGTCCCAAGGTGAATACCTG	1707
QY	1660	CCTTATTTCAGCTTTTCAGAACTAGCCAAAAGAGGAGTTGAAGTGGCTATAGTGACAAT	1719
Db	1708	TCATTATTTCAGCTTTCGCAAACTAGCCAGAAAAGAAATTGAAGGACCTACAGCGACAAC	1767
QY	1720	CCAAATATTGAGATGCTTCTATTGCCAACAGTGGAAAACCATCCACCTCATTAATGCAC	1779
Db	1768	CCCATATCGCCACGCTCCATTGCAAAACAGTGGAAAACCCCTACACCTGATAATGCC	1827
QY	1780	AGTGGAAATGAATGCCACCAATACATTTTATCTCAGTTTCAAAATACAAACGATGAA	1839
Db	1828	GGGGGATGAACGCCACCAAGATCATTTTAACTCTACTTTCAAAACGCCAACACGACGAA	1887
QY	1840	GAGTTTCAAATGCAGATAACAGTGGAGTGGACACAGGGAGGAGCCAAAACATGAATTCT	1899
Db	1888	GAGTTCAAGATCCAGATAGCACTAGAGTGGACACGAGGAGGAGCGCCCAACATGAATTCT	1947
QY	1900	ACGGCCCAAGAGGTTTACGAAATTTAGTTAGTCCCATTAACACTTCTTCAGAGGGGAA	1959
Db	1948	ACAACCCAGAAGGCTTATGAAAGTTTGTTAGCCCACTGACACCTCTTCTCAGGCTGAC	2007
QY	1960	ATCCTTTTTCAGGATATCCCAAGAAAAACGCTTCCCGAAGTTTAAAGACATGATGTT	2019
Db	2008	GTCCCTTTTGAAGATGTCCCAAGAGAAACGCTTCCCAAGATTCAGGACACATGATGA	2067
QY	2020	AACTCAACGAAGGAGCCGAGGAGAGGTGAAAATTCCTGTTGTTAAATATTTTCACCTCCTT	2079
Db	2068	AATGCAACAGGAGATTCGAAGAGGTGAAAATTCCTCGGTTAAATATTTTCACCTCCTT	2127
QY	2080	CCAAAAGACGCCAGTTGAGTCTCAATACCTTGGATTTCGAACATGGAACTGAGACATC	2139
Db	2128	CCAAAAGACGCCAGGTGAGGCTGAGCAACTTGGATTTCGAATTCGAACGTGAGACATC	2187
QY	2140	ACTTTGAAAGGATACAAATTTGTCCAAGTCACGCTTGTGAGATCATTTCTGTGTAACATCA	2199

```

QY 3280 GCATATAAGGACAAACAAATATAGTTTGAATCATGGGAGAGAAATCGCTTTT 3339
Db 3265 GCCTATAAAGACAAACAAATATACAGTTTGAATCATGGGAGAGAAATCGCTTTC 3324
QY 3340 AAAATGATCGTACCAACGTTTCTCATGTGGTGGCCAGTTGGATGACATAAGAAAAAC 3399
Db 3325 AGATGATACGACCAACGTTTCTCATGTGGTGGTGGTGGATGACATCAGAAAAAC 3384
QY 3400 CTTAGGAAGTTGTTTGGCTGAATGACAAACATGACCAACATCAPAAGATGCTCAGACA 3459
Db 3385 CCAGGAAGTCTGTTGTCTGAATGACAAACATGACCAACATCAGGATGCTCCGGACA 3444
QY 3460 GTGAAGGCTGTTCTCAGGACATCTATGAATCCATGTTCCCATACCTTCCCAATTTGAA 3519
Db 3445 GTGAAGGCTGTTCTCAGGACATCTATGAATCCATGTTTCCCATACCTTCCCAATTTGAG 3504
QY 3520 CTGCCAAGAGATGATGAAACGTTTCTTCTCATATGATGATGCTGAGGATGAGAGGCT 3579
Db 3505 CTGCCAAGAGATGATGAAACGTTTCTTCTCATATGATGATGCTGAGGATGAGAGGCT 3564
QY 3580 TATCGAGACAAATTTGAAGT 3598
Db 3565 TATCGAGACAAATTTGAAGT 3583

```

RESULT 9

```

US-09-864-761-10195/c
; Sequence 10195, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408

```

```

; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 10195
; LENGTH: 1962
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005409.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.98
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 0.93
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.4
; US-09-864-761-10195

```

```

Query Match 30.6%; Score 1103.2; DB 10; Length 1962;
Best Local Similarity 99.7%; Pred. No. 5.9e-308;
Matches 1105; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1581 AGATCATTTTCATGAATGTATATAAGTGATCCTTCTCCCAACACGACTCAGCTATATAT 1640
Db 1364 AGATCATTTTCATGAATGTATATAAGTGATCCTTCTCCCAACACGACTCAGCTATATAT 1305
QY 1641 TCCAAAAGGTGAATGCTGCTTATTTTCAGCTTTGCGAAGATAGCCAAAAGAGAGTTGA 1700
Db 1304 TCCAAAAGGTGAATGCTGCTTATTTTCAGCTTTGCGAAGATAGCCAAAAGAGAGTTGA 1245
QY 1701 AGGTGCTCTATAGTGACCAATCAATAATTCGACATGCTTCTATTGCCAACAGTGGAAAC 1760
Db 1244 AGGTGCTCTATAGTGACCAATCAATAATTCGACATGCTTCTATTGCCAACAGTGGAAAC 1185
QY 1761 CATCCAGCTCATATGACAGTGAATGATGCCACCAATCATTTTAACTCTCAGCTT 1820
Db 1184 CATCCAGCTCATATGACAGTGAATGATGCCACCAATCATTTTAACTCTCAGCTT 1125
QY 1821 TCAAAATACAAACGATGAAGAGTTTCAAAATGCAGATAACAGTGGAGTGACACAAAGGA 1880
Db 1124 TCAAAATACAAACGATGAAGAGTTTCAAAATGCAGATAACAGTGGAGTGACACAAAGGA 1065
QY 1881 GGGACCAAACTGAATTTCTAGGCCCCAGAGGGTTAGCAAAATTTAGTTAGTCCATAAC 1940
Db 1064 GGGACCAAACTGAATTTCTAGGCCCCAGAGGGTTAGCAAAATTTAGTTAGTCCATAAC 1005
QY 1941 ACTTCTCCAGAGCGGAAATCCCTTTTGGAGGATATTTCCCAAGAAACGCTTCCCGAA 2000
Db 1004 ACTTCTCCAGAGCGGAAATCCCTTTTGGAGGATATTTCCCAAGAAACGCTTCCCGAA 945
QY 2001 GTTTAAGAGACATGATGTTAACTCAACAAGAGAGAGCCAGAGAGTGAATAATTCCTCT 2060
Db 944 GTTTAAGAGACATGATGTTAACTCAACAAGAGAGAGCCAGAGAGTGAATAATTCCTCT 885
QY 2061 GGTAAATATTTCTACTCTCTCCAAAGAGCGCCAGTTGAGTCTCAATACCTTGGATTGCA 2120
Db 884 GGTAAATATTTCTACTCTCTCCAAAGAGCGCCAGTTGAGTCTCAATACCTTGGATTGCA 825
QY 2121 ACTGGAACATGGACATCACTTTGAAAGGATACAATTTGTCAGTCCAGCTTGTCTGAG 2180
Db 824 ACTGGAACATGGACATCACTTTGAAAGGATACAATTTGTCAGTCCAGCTTGTCTGAG 765
QY 2181 ATCAATTTCTGATGAACCTCACAGCATGCTAAATAAATAAATAAATAAATAAATAAATA 2240
Db 764 ATCAATTTCTGATGAACCTCACAGCATGCTAAATAAATAAATAAATAAATAAATAAATA 705
QY 2241 AACAAATGACAGTTTGGTGGCTCCACAGGAAAAACAGTTTCATATAAAGCATCTTGCCAAA 2300
Db 704 AACAAATGACAGTTTGGTGGCTCCACAGGAAAAACAGTTTCATATAAAGCATCTTGCCAAA 645
QY 2301 CAGCTTAGGAGTGTCTGAAACATTTGACAGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 2360

```

Db 644 CAGCTTAGGAGTGTCTGAAGAGATTGCAGAGGTTGACTTTTCTCGAGTGAGTAAAGT 585  
QY 2361 GAATGTCATGACCGGGTCAGAATCCACCCTCGACTTGGAGACACAGCAAGATTAG 2420  
Db 584 GAATGTCATGACCGGGTCAGAATCCACCCTCGACTTGGAGACACAGCAAGATTAG 525  
QY 2421 AGTGGAACTCACACCCAAAACCATAGCGGAAATGTGACAAAAGAAAAGCCCCCATC 2480  
Db 524 AGTGGAACTCACACCCAAAACCATAGCGGAAATGTGACAAAAGAAAAGCCCCCATC 465  
QY 2481 TCTGATTGTTCCACTGGAAGCCAGATGACAAAAGAAAAGAAAATCACAGGGAAGAAAA 2540  
Db 464 TCTGATTGTTCCACTGGAAGCCAGATGACAAAAGAAAAGAAAATCACAGGGAAGAAAA 405  
QY 2541 AGAAGACAGTAGAATGGAGGAAAATGCTGAAAATCACATAGCGGTTACTGAAGTTTACT 2600  
Db 404 AGAAGACAGTAGAATGGAGGAAAATGCTGAAAATCACATAGCGGTTACTGAAGTTTACT 345  
QY 2601 TGGAGAAAGCTGCAGCATTACACAGATAGTTACTTGGGCTTTTGGCCATGGGAGAAAA 2660  
Db 344 TGGAGAAAGCTGCAGCATTACACAGATAGTTACTTGGGCTTTTGGCCATGGGAGAAAA 285  
QY 2661 AAAGTATTCTCAGATCTCTCGAGAA 2688  
Db 284 AAAGTATTCTCAGATCTCTCGAGGA 257

## RESULT 10

US-09-864-761-26835/c  
; Sequence 26835, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aecmca-x-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21

; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 26835  
; LENGTH: 1125  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC005409.1  
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.5  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.4  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.4  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.98  
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 0.93  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.4  
; OTHER INFORMATION: EST\_HUMAN HIT: A1499228.1, EVALUE 0.00e+00  
; OTHER INFORMATION: NT HIT: AB033034.1, EVALUE 0.00e+00  
; OTHER INFORMATION: SWISSPROT HIT: P36166, EVALUE 3.30e-01  
US-09-864-761-26835

Query Match 30.5%; Score 1101.8; DB 10; Length 1125;  
Best Local Similarity 99.8%; Pred. No. 1e-307;  
Matches 1103; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1581 AGATCATTTTCATCAATTTGTATAAAGTGATCCCTTCTCCCAAAACACACTCATATATAT 1640  
Db 1105 AGATCATTTTCATCAATTTGTATAAAGTGATCCCTTCTCCCAAAACACACTCATATATAT 1046  
QY 1641 TCCAAAAGGTGAATGCTGCTCTTATTTCAGCTTTTCAGAGTAGGCCAAAAGAGGAGTTGA 1700  
Db 1045 TCCAAAAGGTGAATGCTGCTCTTATTTCAGCTTTTCAGAGTAGGCCAAAAGAGGAGTTGA 986  
QY 1701 AGTGGCTATAGTGACAATCCAAATATTCGACATGCTTCTATTGCGCAAGTGGAAAAAC 1760  
Db 985 AGTGGCTATAGTGACAATCCAAATATTCGACATGCTTCTATTGCGCAAGTGGAAAAAC 926  
QY 1761 CATCCACCTCATATAATGCACAGTGAATGAATGCCACCAACAATACATTTTAATCTCACGTT 1820  
Db 925 CATCCACCTCATATAATGCACAGTGAATGAATGCCACCAACAATACATTTTAATCTCACGTT 866  
QY 1821 TCAAAATACAAACGATGAAGAGTTCAAAATGCAGATAACAGTGGAGTGGACACAGGGA 1880  
Db 865 TCAAAATACAAACGATGAAGAGTTCAAAATGCAGATAACAGTGGAGTGGACACAGGGA 806  
QY 1881 GGGACCAAACTGAATTTCTACGGCCCAAGAGGTTACGAAAATTTAGTTAGTCCCATAC 1940  
Db 805 GGGACCAAACTGAATTTCTACAGCCCAAGAGGTTACGAAAATTTAGTTAGTCCCATAC 746  
QY 1941 ACTTCTTCCAGAGCGGAAATCCTTTTTCAGGATATTTCCCAAGAAAACGCTTCCCGAA 2000  
Db 745 ACTTCTTCCAGAGCGGAAATCCTTTTTCAGGATATTTCCCAAGAAAACGCTTCCCGAA 686  
QY 2001 GTTTAAGACATGATGTTTAATCAACAGGAGCCAGGAGAGGTTGAAAATTTCCCTT 2060  
Db 685 GTTTAAGACATGATGTTTAATCAACAGGAGCCAGGAGAGGTTGAAAATTTCCCTT 626  
QY 2061 GGTAAATATTTTCACTCTTCCAAAAGACGCCAGTTGAGTCTCAATACCTTGGATTTGA 2120  
Db 625 GGTAAATATTTTCACTCTTCCAAAAGACGCCAGTTGAGTCTCAATACCTTGGATTTGA 566  
QY 2121 ACTGGAACATGGAGACATCATCTTTGAAAAGGATACAAATTTGCCAAGTACGCTTCTCTGAG 2180  
Db 565 ACTGGAACATGGAGACATCATCTTTGAAAAGGATACAAATTTGCCAAGTACGCTTCTCTGAG 506  
QY 2181 ATCATTTCTGTGTAAGTCAAGCATGCTTAAATAAAAAATCAAGCTATAATAACAGATGA 2240  
Db 505 ATCATTTCTGTGTAAGTCAAGCATGCTTAAATAAAAAATCAAGCTATAATAACAGATGA 446  
QY 2241 AACAATGACAGTTTGGTGGCTTCCACAGAAAACAGGTTTCAATAAACGATCTTGGCCAAA 2300  
|||||



; ORGANISM: Rattus rattus  
US-09-895-072-12

Query Match: 6.5%; Score 233.4; DB 10; Length 460;  
Best Local Similarity 82.7%; Pred. No. 6.3e-57;  
Matches 267; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 666 ATTTCACCAACATTCAGGAGAAACAATCAACATAAATAATTCAGGAGAAATCTTTC 725  
DB 1 ATTCCCAACCAATTCAGGAGAGAGTACGTGAGACACAAACTGCCAGAAAATCTTTC 60  
QY 726 CTCATAAGTCAACTGTTCAGTGTGATTACAGAGGCGAGTGTAGCGCTCTTAAACTGAA 785  
DB 61 TTCTAAATAAATACTGTTCAGCTGTACTCGGAGGCCGCGCTCTTCTGAAATGAA 120  
QY 786 TAACCCCAAGGATTTTCAAGAATTGAATAGCAAACTAGAGAAACATGACCATGATGG 845  
DB 121 TAACCCCAAGGTTTCCCGAGCTGAACAGCAGACCAAGAAACATGAGCATCAGTGG 180  
QY 846 AAAGAACTGACCATAGCTCTCATATTTATGAGGATCTGAGCGCCATCAGCCAGTC 905  
DB 181 GAAGAACTGGCCATCAGCCCTGCTATCTGTGTGGACCTGAGCGCCATCAGCCAGTC 240  
QY 906 TAAGCAGGATGAAGACATCTCTCCAGTGTCTTGAAGATAACGAAGAACTGAGGTACTC 965  
DB 241 CAAGCAGGATGAAGATGTGTCTCCAGCGCTTCGAGGATAACGAAGAGCTGAGGTACTC 300  
QY 966 ATTGCGATCTATCGAGGAGCATG 988  
DB 301 ACTGAGATCTATCGAGAGACATG 323

## RESULT 14

US-09-986-552-12  
; Sequence 12, Application US/09986552  
; Patent No. US20020150981A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
; FILE REFERENCE: 215089US77DIV  
; CURRENT APPLICATION NUMBER: US/09/986,552  
; PRIOR FILING DATE: 2001-11-09  
; PRIOR APPLICATION NUMBER: 09/635,872  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 12  
; LENGTH: 460  
; TYPE: DNA  
; ORGANISM: Rattus rattus  
US-09-986-552-12

Query Match: 6.5%; Score 233.4; DB 10; Length 460;  
Best Local Similarity 82.7%; Pred. No. 6.3e-57;  
Matches 267; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 666 ATTTCACCAACATTCAGGAGAAACAATCAACATAAATAATTCAGGAGAAATCTTTC 725  
DB 1 ATTCCCAACCAATTCAGGAGAGAGTACGTGAGACACAAACTGCCAGAAAATCTTTC 60  
QY 726 CTCATAAGTCAACTGTTCAGTGTGATTACAGAGGCGAGTGTAGCGCTCTTAAACTGAA 785  
DB 61 TTCTAAATAAATACTGTTCAGCTGTACTCGGAGGCCGCGCTCTTCTGAAATGAA 120  
QY 786 TAACCCCAAGGATTTTCAAGAATTGAATAGCAAACTAGAGAAACATGACCATGATGG 845  
DB 121 TAACCCCAAGGTTTCCCGAGCTGAACAGCAGACCAAGAAACATGAGCATCAGTGG 180  
QY 846 AAAGAACTGACCATAGCTCTCATATTTATGAGGATCTGAGCGCCATCAGCCAGTC 905  
DB 181 GAAGAACTGGCCATCAGCCCTGCTATCTGTGTGGACCTGAGCGCCATCAGCCAGTC 240

QY 906 TAAGCAGGATGAAGACATCTCTCCAGTGTCTTTGAAGATAACGAAGAACTGAGGTACTC 965  
DB 241 CAAGCAGGATGAAGATGTGTCTCCAGCGCTTCGAGGATAACGAAGAGCTGAGGTACTC 300  
QY 966 ATTGCGATCTATCGAGAGGAGCATG 988  
DB 301 ACTGAGATCTATCGAGAGACATG 323

## RESULT 15

US-09-864-761-10706/c  
; Sequence 10706, Application us/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aecm1ca-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 10706  
; LENGTH: 459  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC005409.1  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.89  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.68  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.95

; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.93  
US-09-864-761-10706

Query Match	5.8%	Score 208.8;	DB 10;	Length 459;
Best Local Similarity	99.1%	Pred. No. 8.1e-50;		
Matches 210;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;

Qy	334	AGTGAGAGCAGTTAGAGTGTTCCTAACACACTGCATTAAAGGTGCCAATGCTTGCTCG	393
Db	438	AGTGAGAGCAGTTAGAGTGTTCCTAACACACTGCATTAAAGGTGCCAATGCTTGCTCG	379
Qy	394	GACCCAGCCCTGCCAGCCCAACATCACCTGAAGGACCTGCCATCTCTTTATCCTTCTTTT	453
Db	378	GACCCAGCCCTGCCAGCCCAACATCACCTGAAGGACCTGCCATCTCTTTATCCTTCTTTT	319
Qy	454	CATTCTGCCAGTGACATTTTCAATGTTGCCAAACCAAAACCCCTTCTACCAATGTCTCA	513
Db	318	CATTCTGCCAGTGACATTTTCAATGTTGCCAAACCAAAACCCCTTCTACCAATGTCTCA	259
Qy	514	GTTGTGTTTTTTTGACAGTACTAAGGATGTTGA	545
Db	258	GTTGTGTTTTTTTGACAGTACTAAGGATGTTGA	227

Search completed: April 13, 2003, 03:03:45  
Job time : 225 secs





QY 275 KKNWTDGKELTISPAYLLWDLISAISQSKQDEDISASRFEDNEELRYSLRISIERHAPWVR 334  
DB 285 KKNWTDGKELTISPAYLLWDLISAISQSKQDEDISASRFEDNEELRYSLRISIERHAPWVR 344  
QY 335 NFIIVTNGQIPSWLNDNPRVTIVTHQDVFNRLSHLPTFSSPAIESHVHRIEGLSOKFIY 394  
DB 345 NFIIVTNGQIPSWLNDNPRVTIVTHQDVFNRLSHLPTFSSPAIESHVHRIEGLSOKFIY 404  
QY 395 LNDVDMFGKDVMPDDDFYSHSGKQKYLTPVPCNAEGCGPGSWIKDGYCDKACNNSACDWD 454  
DB 405 LNDVDMFGKDVMPDDDFYSHSGKQKYLTPVPCNAEGCGPGSWIKDGYCDKACNNSACDWD 464  
QY 455 GDCSGNSGGRYIAGGCGTSGIGVQPGQGGINSVSYNQCANSWLADKFCDOACN 514  
DB 465 GDCSGNSGGRYIAGGCGTSGIGVQPGQGGINSVSYNQCANSWLADKFCDOACN 524  
QY 515 VLSGCFDAGDCGQDHFHLYKVIILLPNQTHYIIPKGECLPYFSAEVAKRGEVAYSDNP 574  
DB 525 VLSGCFDAGDCGQDHFHLYKVIILLPNQTHYIIPKGECLPYFSAEVAKRGEVAYSDNP 584  
QY 575 IIRHASIAKWKTIHLIMHSGMNATTIHFNLTFQNTDEEFKQITVEVDTREGPKLNST 634  
DB 585 IIRHASIAKWKTIHLIMHSGMNATTIHFNLTFQNTDEEFKQITVEVDTREGPKLNST 644  
QY 635 AOKGYENLVSPITLLPEAEILFEDIPKEKRPKFRHDVNSTRAQAEVKIPIVNIISLLP 694  
DB 645 AOKGYENLVSPITLLPEAEILFEDIPKEKRPKFRHDVNSTRAQAEVKIPIVNIISLLP 704  
QY 695 KDAQLSLNTLDLQLEHGDITLKGYNLSKALLRSFLMNSQAHAKIKNOAITDETNDLSVA 754  
DB 705 KDAQLSLNTLDLQLEHGDITLKGYNLSKALLRSFLMNSQAHAKIKNOAITDETNDLSVA 764  
QY 755 POEKQVHKSILPNSLGVSRERLQRTFFPAVSVKVNHDQGNPPDLLETTARFRVETHQK 814  
DB 765 POEKQVHKSILPNSLGVSRERLQRTFFPAVSVKVNHDQGNPPDLLETTARFRVETHQK 824  
QY 815 TIGGNVTEKPPSLIVPLESOMTEKKITGKEKENSMEENAEHIGVTEVLLGRKLOHY 874  
DB 825 TIGGNVTEKPPSLIVPLESOMTEKKITGKEKENSMEENAEHIGVTEVLLGRKLOHY 884  
QY 875 TDSYLGFLPWEKKYFQDLDDEESLKTQLAYFTDSKNR 915  
DB 885 TDSYLGFLPWEKKYFQDLDDEESLKTQLAYFTDSKNTR 925

## RESULT 2

US-09-986-552-1

; Sequence 1, Application us/09986552

; Patent No. US20020150981A1

; GENERAL INFORMATION:

; APPLICANT: CANFIELD, WILLIAM

; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES

; FILE REFERENCE: 215089US77DIV

; CURRENT APPLICATION NUMBER: US/09/986,552

; CURRENT FILING DATE: 2001-11-09

; PRIOR APPLICATION NUMBER: 09/635,872

; PRIOR FILING DATE: 2000-08-10

; PRIOR FILING DATE: 1999-09-14

; NUMBER OF SEQ ID NOS: 52

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 1

; LENGTH: 928

; TYPE: PR

; ORGANISM: Homo sapiens

US-09-986-552-1

Query Match 73.0%; Score 4630; DB 10; Length 928;

Best Local Similarity 99.3%; Pred. No. 0;

Matches 875; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 35 SRDQYHVLFDSDYRDNIAGKSFQNRCLPMPIDVVTWVNGTDLLELKLQVREQMEEQ 94

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 15  
; LENGTH: 908  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-895-072-15

Query Match 57.4%; Score 3639; DB 10; Length 908;  
Best Local Similarity 78.7%; Pred. No. 5.1e-258;  
Matches 694; Conservative 60; Mismatches 106; Indels 22; Gaps 7;

Qy 35 SRDQYHVLFDSDYRNIAGKSFQNRCLCLPMPIDVVYTWNGTDLLELLKELQOVRHEMEEQ 94  
Db 45 SRDQYHVLFDSDYRNIAGKSFQNRCLCLPMPIDVVYTWNGTDLLELLKELQOVRHEMEEQ 104

Qy 95 KAMREILGKNTTEPTKKSEKOLECLLTHCICKVPMVLDPALPANITLKDLPSPYSEFHAS 154  
Db 105 RAMRETLGKNTTEPTKKSEKOLECLLTHCICKVPMVLDPALPANITLKDLPSPYSEFHAS 164

Qy 155 SDIFNVAKPNPSTNVSVVDFSTKDVEDAHSGLLKGNRSQRTVWRGYLTDDKEVPGGLVLM 214  
Db 165 SDMFNVAKPNPSTNVSVVDFSTKDVEDAHSGLLKGNRSQRTVWRGYLTDDKEVPGGLVLI 224

Qy 215 QDLAFLSGFPPTFKETNQLTKLPEN-LSSKVKLLQLYSEASVALLKNNPKDFOELNKO 273  
Db 225 QGLAFLSGFPPTFKETNQLTKLPEN-LSSKVKLLQLYSEASVALLKNNPKDFOELNKO 284

Qy 274 TKKNTIDGKELTISPAYLLWDLAISQSKODEDISASRFEDNEELRYSLSRTERHAPWY 333  
Db 285 TKKNTIDGKELTISPAYLLWDLAISQSKODEDISASRFEDNEELRYSLSRTERHAPWY 344

Qy 334 RNIFVTNGQIPSWNLNDNPRVTIVTHQDVFNRLSHLPTSSPAIESHVRIEGLSQKFI 393  
Db 345 RNIFVTNGQIPSWNLNDNPRVTIVTHQDVFNRLSHLPTSSPAIESHVRIEGLSQKFI 404

Qy 394 YLNDVDFGKDVWPDFFYSHSGKGVLTWPVPCAGCGPWSWIKDGYCDKACNNSACDW 453  
Db 405 YLNDVDFGKDVWPDFFYSHSGKGVLTWPVPCAGCGPWSWIKDGYCDKACNNSACDW 464

Qy 454 DGGDCSGNNGSRYIAGGGTSGVGPWFQGGGINSYCYNOGCANSWLADKFCDOAC 513  
Db 465 DGGDCSGNNGSRYIAGGGTSGVGPWFQGGGINSYCYNOGCANSWLADKFCDOAC 524

Qy 514 NVLSCGFDAGCGQDHFHELYKVTLLPNQTHYIIPKGECLPYFSFAEVAKRVEGAYSDN 573  
Db 525 NVLSCGFDAGCGQDHFHELYKVTLLPNQTHYIIPKGECLPYFSFAEVAKRVEGAYSDN 584

Qy 574 PIIRHASIANKWKTILHIMHSGMNATTIHFNLTFQNTDEEFKMOITVEVDTRREGPKLNS 633  
Db 585 PIIRHASIANKWKTILHIMHSGMNATTIHFNLTFQNTDEEFKMOITVEVDTRREGPKLNS 644

Qy 634 TAQKGYENLVSPITLLPEAEILFEDIPKRRFPKRRHDVNSTRAAQEVEKIPLVNISLL 693  
Db 645 TTQKAYESLVSPVTPLOADVFEDVPKRRFPKRRHDVNSTRAAQEVEKIPLVNISLL 704

Qy 694 PKDAQSLNTLDLQLEHGDITLKGYNLSKALLRSFLMNSHAKIKNOAIITDETNDLSV 753  
Db 705 PKDAQSLNTLDLQLEHGDITLKGYNLSKALLRSFLMNSHAKIKNOAIITDETNDLSV 763

Qy 754 APQEKQVHKSILPNSLGVSRLOLTPPAVSKVNGHDGQGNPDLETTARFRVETHTQ 813  
Db 764 VPQENPSHRR--PHGFAGEHRSERWTAPAEITVTKGRDHALNPPVLETWNL 816

Qy 814 KTIGGNVTKPPSLIVPLESQMTEKKITGKEKENSMEENAEHNIGVTEVLGLRKLQH 873  
Db 817 PTLGTVTSKENLSPLIVPPESHLP-----KEEDSRAEGNA---VPVKELVPGRRLQ- 865

Qy 874 YTDSVLGLPWEKKYFLLDDEESLKTQLAYETDSEKNR 915  
Db 866 --QNYPGFLPWEKKYFLLDDEESLKTQLAYETDSEKNR 905

RESULT 4  
US-09-986-552-15  
; Sequence 15, Application US/09986552  
; Patent No. US20020150981A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
; FILE REFERENCE: 215089US77DIV  
; CURRENT APPLICATION NUMBER: US/09/986,552  
; PRIOR FILING DATE: 2001-11-09  
; PRIOR APPLICATION NUMBER: 09/635,872  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 15  
; LENGTH: 908  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-986-552-15

Query Match 57.4%; Score 3639; DB 10; Length 908;  
Best Local Similarity 78.7%; Pred. No. 5.1e-258;  
Matches 694; Conservative 60; Mismatches 106; Indels 22; Gaps 7;

Qy 35 SRDQYHVLFDSDYRNIAGKSFQNRCLCLPMPIDVVYTWNGTDLLELLKELQOVRHEMEEQ 94  
Db 45 SRDQYHVLFDSDYRNIAGKSFQNRCLCLPMPIDVVYTWNGTDLLELLKELQOVRHEMEEQ 104

Qy 95 KAMREILGKNTTEPTKKSEKOLECLLTHCICKVPMVLDPALPANITLKDLPSPYSEFHAS 154  
Db 105 RAMRETLGKNTTEPTKKSEKOLECLLTHCICKVPMVLDPALPANITLKDLPSPYSEFHAS 164

Qy 155 SDIFNVAKPNPSTNVSVVDFSTKDVEDAHSGLLKGNRSQRTVWRGYLTDDKEVPGGLVLM 214  
Db 165 SDMFNVAKPNPSTNVSVVDFSTKDVEDAHSGLLKGNRSQRTVWRGYLTDDKEVPGGLVLI 224

Qy 215 QDLAFLSGFPPTFKETNQLTKLPEN-LSSKVKLLQLYSEASVALLKNNPKDFOELNKO 273  
Db 225 QGLAFLSGFPPTFKETNQLTKLPEN-LSSKVKLLQLYSEASVALLKNNPKDFOELNKO 284

Qy 274 TKKNTIDGKELTISPAYLLWDLAISQSKODEDISASRFEDNEELRYSLSRTERHAPWY 333  
Db 285 TKKNTIDGKELTISPAYLLWDLAISQSKODEDISASRFEDNEELRYSLSRTERHAPWY 344

Qy 334 RNIFVTNGQIPSWNLNDNPRVTIVTHQDVFNRLSHLPTSSPAIESHVRIEGLSQKFI 393  
Db 345 RNIFVTNGQIPSWNLNDNPRVTIVTHQDVFNRLSHLPTSSPAIESHVRIEGLSQKFI 404

Qy 394 YLNDVDFGKDVWPDFFYSHSGKGVLTWPVPCAGCGPWSWIKDGYCDKACNNSACDW 453  
Db 405 YLNDVDFGKDVWPDFFYSHSGKGVLTWPVPCAGCGPWSWIKDGYCDKACNNSACDW 464

Qy 454 DGGDCSGNNGSRYIAGGGTSGVGPWFQGGGINSYCYNOGCANSWLADKFCDOAC 513  
Db 465 DGGDCSGNNGSRYIAGGGTSGVGPWFQGGGINSYCYNOGCANSWLADKFCDOAC 524

Qy 514 NVLSCGFDAGCGQDHFHELYKVTLLPNQTHYIIPKGECLPYFSFAEVAKRVEGAYSDN 573  
Db 525 NVLSCGFDAGCGQDHFHELYKVTLLPNQTHYIIPKGECLPYFSFAEVAKRVEGAYSDN 584

Qy 574 PIIRHASIANKWKTILHIMHSGMNATTIHFNLTFQNTDEEFKMOITVEVDTRREGPKLNS 633  
Db 585 PIIRHASIANKWKTILHIMHSGMNATTIHFNLTFQNTDEEFKMOITVEVDTRREGPKLNS 644

Qy 634 TAQKGYENLVSPITLLPEAEILFEDIPKRRFPKRRHDVNSTRAAQEVEKIPLVNISLL 693  
Db 645 TTQKAYESLVSPVTPLOADVFEDVPKRRFPKRRHDVNSTRAAQEVEKIPLVNISLL 704

Qy 694 PKDAQSLNTLDLQLEHGDITLKGYNLSKALLRSFLMNSHAKIKNOAIITDETNDLSV 753  
Db 705 PKDAQSLNTLDLQLEHGDITLKGYNLSKALLRSFLMNSHAKIKNOAIITDETNDLSV 763

Qy	754	APQEKQVHKSIPLNSLGVSERLQRLTFPAVSVKVNHGDOQNPPDLTETARFVETHQ	813
Db	764	VPQENPSHRR--PHGFAGEHRSERWAPAEFTVTKGRDHALNPPPVLETNARL----	816
Qy	814	KTIGNNVTKPKPSLPIVPLESQMTKEKKTIGKENSMEENAEHIGVTEVLLGRKLQH	873
Db	817	PTLGVTVSKENSLPIVPPESHLP-----KKEEDSRAEGNA--VPKELVPGERLQ-	865
Qy	874	YTSYLGFLPWEKKKYFLDLDDEESLKTQLAYFTDSKNRAR	915
Db	866	--ONYGFLPWEKKKYFQDLDDEESLKTQLAYFTDRKHTGR	905

RESULT 5  
US-09-864-761-42893  
; Sequence 42893, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aeomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-03-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 2463.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 42893  
; LENGTH: 367  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC005409.1  
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.5

Db 61 EDMQAFSYFYLLMSAVQPLNISQVFEVDVTDQSGVLSDRIRTLATRIHELPLSLQDLT 120  
QY 1039 GLEHMLNCSKMLPADITQLNNIPPTQESYDDPNLPPVTKSLVTNCKPVTDKIHKAYKDK 1098  
Db 121 GLEHMLNCSKMLPADITQLNNIPPTQESYDDPNLPPVTKSLVTNCKPVTDKIHKAYKDK 180  
QY 1099 NKYRFEIMGEEIEAFKMITNTVSHVVGQDLDIRKPNKRVCLNDNIDHNDHKAQTVKAVL 1158  
Db 181 NKYRFEIMGEEIEAFKMITNTVSHVVGQDLDIRKPNKRVCLNDNIDHNDHKAQTVKAVL 240  
QY 1159 RQFYESMFPPIPSQFELPREYRNRLHMHLEQWRAIRYDKLK 1199  
Db 241 RQFYESMFPPIPSQFELPREYRNRLHMHLEQWRAIRYDKLK 281

RESULT 7  
US-09-986-552-2  
; Sequence 2, Application US/09986552  
; Patent No. US20020150981A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
; FILE REFERENCE: 215089US77DIV  
; CURRENT APPLICATION NUMBER: US/09/986,552  
; PRIOR FILING DATE: 2001-11-09  
; PRIOR APPLICATION NUMBER: 09/635,872  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 2  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-986-552-2

Query Match 23.5%; Score 1487; DB 10; Length 328;  
Best Local Similarity 100.0%; Pred. No. 5.1e-101;  
Matches 281; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 919 DTFADSLRYVKNILNSKFGFTSRKVPAPAHMPHMDIRIVMQELQDMFPEEFDKTSFHKVRRHS 978  
Db 1 DTFADSLRYVKNILNSKFGFTSRKVPAPAHMPHMDIRIVMQELQDMFPEEFDKTSFHKVRRHS 60  
QY 979 EDMQAFSYFYLLMSAVQPLNISQVFEVDVTDQSGVLSDRIRTLATRIHELPLSLQDLT 1038  
Db 61 EDMQAFSYFYLLMSAVQPLNISQVFEVDVTDQSGVLSDRIRTLATRIHELPLSLQDLT 120  
QY 1039 GLEHMLNCSKMLPADITQLNNIPPTQESYDDPNLPPVTKSLVTNCKPVTDKIHKAYKDK 1098  
Db 121 GLEHMLNCSKMLPADITQLNNIPPTQESYDDPNLPPVTKSLVTNCKPVTDKIHKAYKDK 180  
QY 1099 NKYRFEIMGEEIEAFKMITNTVSHVVGQDLDIRKPNKRVCLNDNIDHNDHKAQTVKAVL 1158  
Db 181 NKYRFEIMGEEIEAFKMITNTVSHVVGQDLDIRKPNKRVCLNDNIDHNDHKAQTVKAVL 240  
QY 1159 RQFYESMFPPIPSQFELPREYRNRLHMHLEQWRAIRYDKLK 1199  
Db 241 RQFYESMFPPIPSQFELPREYRNRLHMHLEQWRAIRYDKLK 281

RESULT 8  
US-09-895-072-8  
; Sequence 8, Application US/09895072  
; Patent No. US2002025550A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM M  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
; FILE REFERENCE: 210119USOCONT  
; CURRENT APPLICATION NUMBER: US/09/895,072  
; CURRENT FILING DATE: 2001-07-02

; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; PRIOR APPLICATION NUMBER: US 09/635,872  
; PRIOR FILING DATE: 2000-08-10  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 8  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-895-072-8

Query Match 23.1%; Score 1465; DB 10; Length 328;  
Best Local Similarity 98.2%; Pred. No. 2.1e-99;  
Matches 276; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
QY 919 DTFADSLRYVKNILNSKFGFTSRKVPAPAHMPHMDIRIVMQELQDMFPEEFDKTSFHKVRRHS 978  
Db 1 DTFADSLRYVKNILNSKFGFTSRKVPAPAHMPHMDIRIVMQELQDMFPEEFDKTSFHKVRRHS 60  
QY 979 EDMQAFSYFYLLMSAVQPLNISQVFEVDVTDQSGVLSDRIRTLATRIHELPLSLQDLT 1038  
Db 61 EDMQAFSYFYLLMSAVQPLNISQVFEVDVTDQSGVLSDRIRTLATRIHELPLSLQDLT 120  
QY 1039 GLEHMLNCSKMLPADITQLNNIPPTQESYDDPNLPPVTKSLVTNCKPVTDKIHKAYKDK 1098  
Db 121 GLEHMLNCSKMLPADITQLNNIPPTQESYDDPNLPPVTKSLVTNCKPVTDKIHKAYKDK 180  
QY 1099 NKYRFEIMGEEIEAFKMITNTVSHVVGQDLDIRKPNKRVCLNDNIDHNDHKAQTVKAVL 1158  
Db 181 NKYRFEIMGEEIEAFKMITNTVSHVVGQDLDIRKPNKRVCLNDNIDHNDHKAQTVKAVL 240  
QY 1159 RQFYESMFPPIPSQFELPREYRNRLHMHLEQWRAIRYDKLK 1199  
Db 241 RQFYESMFPPIPSQFELPREYRNRLHMHLEQWRAIRYDKLK 281

RESULT 9  
US-09-986-552-8  
; Sequence 8, Application US/09986552  
; Patent No. US20020150981A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
; FILE REFERENCE: 215089US77DIV  
; CURRENT APPLICATION NUMBER: US/09/986,552  
; PRIOR FILING DATE: 2001-11-09  
; PRIOR APPLICATION NUMBER: 09/635,872  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/153,831  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 8  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-986-552-8

Query Match 23.1%; Score 1465; DB 10; Length 328;  
Best Local Similarity 98.2%; Pred. No. 2.1e-99;  
Matches 276; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
QY 919 DTFADSLRYVKNILNSKFGFTSRKVPAPAHMPHMDIRIVMQELQDMFPEEFDKTSFHKVRRHS 978  
Db 1 DTFADSLRYVKNILNSKFGFTSRKVPAPAHMPHMDIRIVMQELQDMFPEEFDKTSFHKVRRHS 60  
QY 979 EDMQAFSYFYLLMSAVQPLNISQVFEVDVTDQSGVLSDRIRTLATRIHELPLSLQDLT 1038  
Db 61 EDMQAFSYFYLLMSAVQPLNISQVFEVDVTDQSGVLSDRIRTLATRIHELPLSLQDLT 120  
QY 1039 GLEHMLNCSKMLPADITQLNNIPPTQESYDDPNLPPVTKSLVTNCKPVTDKIHKAYKDK 1098  
Db 241 RQFYESMFPPIPSQFELPREYRNRLHMHLEQWRAIRYDKLK 281

Db 121 GLEHMLNCSKMLPANITQLNPIPTQAYYDNLPPVTKSLVNTCKVPVTDKTHKAYKDK 180  
QY 1099 NKYRPEIMGEEIEAFKMLRTNVSHVVGQDDIRKNPKRFVCLNDNIDHNHKKDAQTVKAVL 1158  
Db 181 NKYRPEIMGEEIEAFKMLRTNVSHVVGQDDIRKNPKRFVCLNDNIDHNHKKDAQTVKAVL 240  
QY 1159 RDFSMEFPPIPSQFELPREYRNRLHMHLEQEWAYRDKLK 1199  
Db 241 RDFSMEFPPIPSQFELPREYRNRLHMHLEQEWAYRDKLK 281  
RESULT 10  
US-09-895-072-13  
; Sequence 13, Application US/09895072  
; Patent No. US2002025550A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM M  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
; FILE REFERENCE: 21019USOCNT  
; CURRENT APPLICATION NUMBER: US/09/895,072  
; CURRENT FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; PRIOR APPLICATION NUMBER: US 09/635,872  
; PRIOR FILING DATE: 2000-08-10  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 13  
; LENGTH: 502  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
US-09-895-072-13

Query Match 11.7%; Score 742; DB 10; Length 502;  
Best Local Similarity 23.2%; Pred. No. 3.8e-46;  
Matches 197; Conservative 98; Mismatches 181; Indels 372; Gaps 14;  
QY 312 RFEDNEELRYSLRSIERHAPVWRNIFVTNGQIPSWMLNDNPRVTIVTHQDVFNLSHLP 371  
Db 4 RFDDKNELRYSLRSLEKHAATIRHVIYVTNGQIPSWLDLSYERVTVPHEVLAPDQLP 63  
QY 372 TFSAPAESHVHRIEGLSOKFIYLNDDVMFGKDVDPDDFYSHSKGOKVYLTWPVPCAE 431  
Db 64 TFSAAIETFLHRIPKLSRFLYLNDDIFLGAPLYPEDLYTEAGVRYQAWVPGCCALD 123  
QY 432 CPGSWIKDGYCDKACNNSACDWDGDCSGNSGGSRITAGGGGTGSGVGPWFQGGINS 491  
Db 124 CPWTYIGDGCADRHCHNIDACQFDGDCSETGPAS 157  
QY 492 VSYCNOGCANSWLADKFCQACNVLSGCFDAGDCQDHFHLYKVILLPNQTHYIIPKGE 551  
Db 158 -----DAHVIPPFSKE 167  
QY 552 CLPYFSFAEVAKRGVAGYSDNPIIRHASIANKWKTIIHLMHSGMNATTIHFNLTFQNTN 611  
Db 168 VLE----- 170  
QY 612 DEEFKMQITVEVDTRGPKLNSTAKGYENLYSPITLLPEAEILFEDIPKRRFPK 667  
Db 171 -----VQP-AAVPQSRV-----HREFQMGQLQ 190  
QY 668 --FKRHDVNSTRAQEEVKIPLVNLISLIPKDAQLSNLTDLQLEHGDITLKGYNLSKAL 725  
Db 191 KLFRRSSANF-----KD-----VMRHRNVSTLKE 214  
QY 726 LRSFLMNSOHAKIKNOAITTDETNDSLVAPQEKOVHKSILNLSGVSERLQRLTFPAVS 785  
Db 215 LRIVRERFNKALMS--LNPLETSSSEPTQTH----- 247  
QY 786 KVNHDQGNQPNPLDLETFARFRVETHQKTIGGNVTKEKPPSLIVPLESQMTKEKKTGK 845  
Db 248 -----GL 249

QY 846 EKENSRENEAENHIGVTEVLLGRKLOHYTDSYLGFLPWEKKKKYFLDILLDEESLKTQIA 905  
Db 250 RKE----- 252  
QY 906 YFTDSKNRARIKRDTPADSLRYVNVKILNSKFGTGRKYPAHMPHMDIRVMQELQDMPE 965  
Db 253 ---DFKS---STDYIYSHLIATNMLNRAYGFKARHVLAVHGVFLDKRDIVAMORRHHQ 305  
QY 966 EFDKTSFHKVRHSEDMQFAFSYFYILMSAVQPLNISQVDFEVDYDQSGVLSREIRTLAT 1025  
Db 306 QILDTAHOFRAPTDLQYAFAYFSLMSETKVMSEIEFDEFDTDGSATWSREVRETLT 365  
QY 1026 RIHELPLSLQDLTGLEHMLNCSK----MLPADITQLNINPTQESYDYPNLPVTKSLV 1081  
Db 366 RIYQPLDWSAMRYFEVYVQNCNRLNGLMKVDVTVEHSTL--VYERYEDSNLPTITRDVL 423  
QY 1082 TCKPVTDKTHKAYDKNRYRFEIMGE--EETAFKMITRNVSHVVGQDDIRKNPKRFVC 1139  
Db 424 VRCPLLAELAANFAVRPKYNFHVSPKTSHSNFMMLISNLTVEVYESLDRLRNRKNC 483  
QY 1140 LNDNIDHN 1147  
Db 484 INDNLDAN 491  
RESULT 11  
US-09-986-552-13  
; Sequence 13, Application US/09986552  
; Patent No. US20020150981A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
; FILE REFERENCE: 215089US77DIV  
; CURRENT APPLICATION NUMBER: US/09/986,552  
; CURRENT FILING DATE: 2001-11-09  
; PRIOR APPLICATION NUMBER: 09/635,872  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 13  
; LENGTH: 502  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
US-09-986-552-13

Query Match 11.7%; Score 742; DB 10; Length 502;  
Best Local Similarity 23.2%; Pred. No. 3.8e-46;  
Matches 197; Conservative 98; Mismatches 181; Indels 372; Gaps 14;  
QY 312 RFEDNEELRYSLRSIERHAPVWRNIFVTNGQIPSWMLNDNPRVTIVTHQDVFNLSHLP 371  
Db 4 RFDDKNELRYSLRSLEKHAATIRHVIYVTNGQIPSWLDLSYERVTVPHEVLAPDQLP 63  
QY 372 TFSAPAESHVHRIEGLSOKFIYLNDDVMFGKDVDPDDFYSHSKGOKVYLTWPVPCAE 431  
Db 64 TFSAAIETFLHRIPKLSRFLYLNDDIFLGAPLYPEDLYTEAGVRYQAWVPGCCALD 123  
QY 432 CPGSWIKDGYCDKACNNSACDWDGDCSGNSGGSRITAGGGGTGSGVGPWFQGGINS 491  
Db 124 CPWTYIGDGCADRHCHNIDACQFDGDCSETGPAS 157  
QY 492 VSYCNOGCANSWLADKFCQACNVLSGCFDAGDCQDHFHLYKVILLPNQTHYIIPKGE 551  
Db 158 -----DAHVIPPFSKE 167  
QY 552 CLPYFSFAEVAKRGVAGYSDNPIIRHASIANKWKTIIHLMHSGMNATTIHFNLTFQNTN 611  
Db 168 VLE----- 170  
QY 612 DEEFKMQITVEVDTRGPKLNSTAKGYENLYSPITLLPEAEILFEDIPKRRFPK 667



Db 171 -----VQP-AAVQSRV-----HRFPQGLQ 190  
QY 668 --FKRHDVNSTRAQEEVKIPLVNISSLPKDAQLSLNTLDLQLEHGDITLKGYNLKSAL 725  
Db 191 KLFRRSANF-----KD-----VMHRNVSTLKE 214  
QY 726 LRSFLMSQAKIKNOAIITDETNDSLVAPQEQVHKSLPNSLGVSRQLRQLTFPAVS 785  
Db 215 LRRIVERENKAKLMS---LNPELETSSSEPQTQRH-----247  
QY 786 KVNCHDQOGNPPLDLETTARFRVETHQTIGGNVTKEKPPSLIVPLESQMTKEKKITGK 845  
Db 248 -----GL 249  
QY 846 EKENSMEENAHNIGVTEVLLGRKLOHYTDSYGLFPLWEKKYFLDLLDEESLKTOLA 905  
Db 250 RKE-----252  
QY 906 YFTDSKNRARKDFTFADSLRYVKNILSKFGFTSRKVPAAHPMHIDRIVMOELQMPFE 965  
Db 253 ---DFKS---STDYSHSLIATNMLNRAYGFKARHVLAVHGVFLDKDIVEAMQRRHQ 305  
QY 966 EFDKTSFHKVRUSEDQMFAPSFYFYLMSAVQPLNISQVDEVDTSQVLSDEIRTIAT 1025  
Db 306 QILDTAQRFRAPTDLDQYAFAYISFLMSETKVSVEEIPDEFDTGCSATWSDEVRITLT 365  
QY 1026 RIHELPLSLQDLTGLFHLMLNSK---MLPADITOLNINPTQESYDPPNLPVTKSLV 1081  
Db 366 RIYQPLDSAMRYFEVQVQCNTRNLGMHLKVDTVEHSTL--VYBRYEDSNLPTITRDIV 423  
QY 1082 TNCKPVTDKIHKAYDKNRYFEIMGE--EEIAFKMIRTNVSHVYGOLDIRKPNRKFVC 1139  
Db 424 VRCPLAALAAANFAVRPKYHVSFKRTHSHSNFMMLISNLTVEVESLDRLRNRKPKNC 483  
QY 1140 LNDNIDHN 1147  
Db 484 INDNLDAN 491

## RESULT 12

US-09-895-072-11

; Sequence 11, Application US/09895072

; Patent No. US20020025550A1

; GENERAL INFORMATION:

; APPLICANT: CANFIELD, WILLIAM M

; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYOSOMAL HYDROLASES

; FILE REFERENCE: 21019USOCNT

; CURRENT APPLICATION NUMBER: US/09/895,072

; CURRENT FILING DATE: 2001-07-02

; PRIOR APPLICATION NUMBER: 60/153,831

; PRIOR FILING DATE: 1999-09-14

; PRIOR APPLICATION NUMBER: US 09/635,872

; PRIOR FILING DATE: 2000-08-10

; NUMBER OF SEQ ID NOS: 52

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 11

; LENGTH: 113

; TYPE: PRT

; ORGANISM: Rattus rattus

US-09-895-072-11

Query Match 7.9%; Score 503; DB 10; Length 113;  
Best Local Similarity 92.5%; Pred. No. 1.2e-29;  
Matches 99; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 223 FPPTFKETNQLTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKTQKNMTIDG 282

Db 1 FPPTFKETSQLTKLPENLSSKIKLLQLYSEASVALLKLNPKGFPPELNKQTKKNMSISG 60

QY 283 KELTSIPAYLLWDLISAISQKODEDISASRFEDNEELRSLRSIERH 329

Db 61 KELTSIPAYLLWDLISAISQKODEDVASRFEDNEELRSLRSIERH 107

## RESULT 13

US-09-986-552-11

; Sequence 11, Application US/09986552

; Patent No. US20020150981A1

; GENERAL INFORMATION:

; APPLICANT: CANFIELD, William

; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYOSOMAL HYDROLASES

; FILE REFERENCE: 215089US77DIV

; CURRENT APPLICATION NUMBER: US/09/986,552

; CURRENT FILING DATE: 2001-11-09

; PRIOR APPLICATION NUMBER: 09/635,872

; PRIOR FILING DATE: 2000-08-10

; PRIOR APPLICATION NUMBER: 60/153,831

; PRIOR FILING DATE: 1999-09-14

; NUMBER OF SEQ ID NOS: 52

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 11

; LENGTH: 113

; TYPE: PRT

; ORGANISM: Rattus rattus

US-09-986-552-11

Query Match 7.9%; Score 503; DB 10; Length 113;

Best Local Similarity 92.5%; Pred. No. 1.2e-29;

Matches 99; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 223 FPPTFKETNQLTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKTQKNMTIDG 282

Db 1 FPPTFKETSQLTKLPENLSSKIKLLQLYSEASVALLKLNPKGFPPELNKQTKKNMSISG 60

QY 283 KELTSIPAYLLWDLISAISQKODEDISASRFEDNEELRSLRSIERH 329

Db 61 KELTSIPAYLLWDLISAISQKODEDVASRFEDNEELRSLRSIERH 107

## RESULT 14

US-10-079-623-349

; Sequence 349, Application US/10079623

; Patent No. US20020169302A1

; GENERAL INFORMATION:

; APPLICANT: Havukkala, Ilkka J.

; APPLICANT: Glenn, Matthew

; APPLICANT: Grigor, Murray R.

; APPLICANT: Molenaar, Adrian J.

; TITLE OF INVENTION: Compositions isolated from bovine

; FILE REFERENCE: 11000,1044c3

; CURRENT APPLICATION NUMBER: US/10/079,623

; CURRENT FILING DATE: 2002-02-19

; NUMBER OF SEQ ID NOS: 370

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 349

; LENGTH: 112

; TYPE: PRT

; ORGANISM: Bovine

US-10-079-623-349

Query Match 7.7%; Score 487; DB 9; Length 112;  
Best Local Similarity 83.9%; Pred. No. 1.8e-28;  
Matches 94; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

QY 851 RMEENAHNIGVTEVLLGRKLOHYTDSYGLFPLWEKKYFLDLLDEESLKTOLAYFTDS 910

Db 1 KIEENANSYPCGNEVPCRKLOQYTDYGLFPLWEKKYFODLLDEESLKTOLAYFTDS 60

QY 911 KNARYKRDFTFADSLRYVKNILSKFGFTSRKVPAAHPMHIDRIVMOELQDM 962

Db 61 KHTGRQLKDTFADSLRYVKNILSKFGFTSRKVPAAHPMHIDRIVMOELQDM 112

## RESULT 15

US-09-864-761-43413

```
; Sequence 43413, Application us/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Acomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 43413
; LENGTH: 68
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005409.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.89
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.68
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.95
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.93
; OTHER INFORMATION: SWISSPROT HIT: P28976, EVALUATE 3.00e-01
US-09-864-761-43413
```

Query Match 5.6%; Score 352; DB 10; Length 68;  
Best Local Similarity 100.0%; Pred. No. 6.5e-19;  
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 113 EKQLECLLTHCIKYPMLVLPALPANITLKDPLSLYPSFHSDIFNVAKPKNPSTNSV 172
Db 1 EKQLECLLTHCIKYPMLVLPALPANITLKDPLSLYPSFHSDIFNVAKPKNPSTNSV 60
```

Qy 173 VVFDSTKD 180  
|||||||  
Db 61 VVFDSTKD 68

Search completed: April 13, 2003, 03:31:00  
Job time : 37.8869 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2003, 03:00:01 ; Search time 0.212801 Seconds  
(without alignments)  
3873.042 Million cell updates/sec

Title: US-10-023-888-22

Perfect score: 20

Sequence: 1 IEGR 4

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_21.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rhodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_rvirus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	41	16 Q8VKR0	Q8VKR0 mycobacteri
2	20	100.0	50	10 Q9LUZ2	Q9LUZ2 arabidopsis
3	20	100.0	55	9 Q9ZX12	Q9ZX12 mycobacteri
4	20	100.0	56	5 Q26922	Q26922 trypanosoma
5	20	100.0	58	4 Q9UPG5	Q9UPG5 homo sapien
6	20	100.0	60	2 Q93U42	Q93U42 carsonella
7	20	100.0	60	2 Q93U40	Q93U40 carsonella
8	20	100.0	61	12 Q91GA0	Q91GA0 columbid ci
9	20	100.0	63	2 Q93FW0	Q93FW0 pseudomonas
10	20	100.0	66	2 Q9AP90	Q9AP90 uncultured
11	20	100.0	66	16 Q51760	Q51760 borrelia bu
12	20	100.0	67	16 Q8YJG9	Q8YJG9 brucella me
13	20	100.0	68	9 Q38263	Q38263 lactococcus
14	20	100.0	70	16 Q8XRG1	Q8XRG1 ralstonia s
15	20	100.0	71	16 Q9PDH9	Q9PDH9 xyella fas
16	20	100.0	73	10 P93514	P93514 gossypium b

17	20	100.0	74	2 Q47817	Q47817 enterococcu
18	20	100.0	75	16 Q8YSR5	Q8YSR5 anabaena sp
19	20	100.0	75	16 Q8YM57	Q8YM57 anabaena sp
20	20	100.0	75	16 Q8XNX3	Q8XNX3 clostridium
21	20	100.0	77	3 Q9P5V0	Q9P5V0 neurospora
22	20	100.0	78	2 P82548	P82548 streptococc
23	20	100.0	78	9 Q37833	Q37833 bacterioph
24	20	100.0	79	11 P97446	P97446 mus musculu
25	20	100.0	80	16 Q8YXU1	Q8YXU1 anabaena sp
26	20	100.0	81	12 Q914G2	Q914G2 sulfolobus
27	20	100.0	82	10 Q80460	Q80460 arabidopsis
28	20	100.0	85	17 Q803Y8	Q803Y8 pyrococcus
29	20	100.0	86	10 Q9MOC7	Q9MOC7 arabidopsis
30	20	100.0	86	12 Q67406	Q67406 influenzavi
31	20	100.0	86	12 Q67407	Q67407 influenzavi
32	20	100.0	86	12 Q67408	Q67408 influenzavi
33	20	100.0	87	6 Q77640	Q77640 macaca mula
34	20	100.0	87	12 Q91U03	Q91U03 influenza a
35	20	100.0	88	6 Q9BDH8	Q9BDH8 bos taurus
36	20	100.0	88	10 Q9ZV45	Q9ZV45 arabidopsis
37	20	100.0	90	10 Q9LFW8	Q9LFW8 arabidopsis
38	20	100.0	91	12 Q91U02	Q91U02 influenza a
39	20	100.0	92	10 Q9ASJ2	Q9ASJ2 oryza sativ
40	20	100.0	92	17 Q97X43	Q97X43 sulfolobus
41	20	100.0	94	10 Q9FTL0	Q9FTL0 oryza sativ
42	20	100.0	94	12 Q8V6R1	Q8V6R1 halovirus h
43	20	100.0	94	16 Q914Y9	Q914Y9 pseudomonas
44	20	100.0	99	15 Q40235	Q40235 human immun
45	20	100.0	99	15 Q40238	Q40238 human immun

#### ALIGNMENTS

#### RESULT 1

Q8VKR0 ID Q8VKR0 PRELIMINARY; PRT; 41 AA.  
AC Q8VKR0; 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Hypothetical protein MT0159.  
GN MT0159.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDC 1551 / OSHKOSH;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains."  
RL Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AE006927; AAK44383.1; -  
DR TIGR; MT0159; -  
KW Hypothetical protein.  
SQ SEQUENCE 41 AA; 4434 MW; 4E1DEA437CAA07B9 CRC64;

Query Match 100.0%; Score 20; DB 16; Length 41;

Best Local Similarity 100.0%; Pred. No. 3.9e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4

Db 26 IEGR 29

#### RESULT 2

Q9LUZ2  
ID Q9LUZ2 PRELIMINARY; PRT; 50 AA.  
AC Q9LUZ2  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE Genomic DNA, chromosome 5, p1 clone:M2N1.  
OS Arabidopsis thaliana (Mouse-ear cross).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RX MEDLINE=20181125; PubMed=10718197;  
RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,  
RA Tabata S.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence  
RT features of the regions of 3,076,755 bp covered by sixty p1 and TAC  
RT clones.";  
RL DNA Res. 7:31-63(2000).  
DR EMBL; AB020755; BAA97334.1; -.  
SQ SEQUENCE 50 AA; 5958 MW; EDEFF7A687643966 CRC64;  
Query Match 100.0%; Score 20; DB 10; Length 50;  
Best Local Similarity 100.0%; Pred. No. 4.8e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 IEGR 4  
Db 39 IEGR 42  
RESULT 3  
Q9ZX12  
ID Q9ZX12 PRELIMINARY; PRT; 55 AA.  
AC Q9ZX12;  
DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)  
DE Gp68.  
OS Mycobacteriophage TM4.  
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.  
OX NCBI\_TaxID=88870;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20110038; PubMed=10645443;  
RA Ford M.E., Stenstrom C., Hendrix R.W., Hatfull G.F.;  
RT "Mycobacteriophage TM4: Genome structure and gene expression.";  
RL Tuber. Lung Dis. 79:63-73(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Ford M.E., Stenstrom C., Hendrix R.W., Hatfull G.F.;  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF068845; AAD17633.1; -.  
SQ SEQUENCE 55 AA; 5782 MW; F2746364DB90A01F CRC64;  
Query Match 100.0%; Score 20; DB 9; Length 55;  
Best Local Similarity 100.0%; Pred. No. 5.3e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 IEGR 4  
Db 51 IEGR 54  
RESULT 4  
Q26922  
ID Q26922 PRELIMINARY; PRT; 56 AA.  
AC Q26922;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE G protein (Fragment).  
OS Trypanosoma cruzi.  
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
OX NCBI\_TaxID=5693;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Y;  
RA Lopes U.G.;  
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; M92652; AAA91960.1; -.  
DR InterPro; IPR001806; Ras\_trnsmfmg.  
DR InterPro; IPR003575; Small\_GTPase.  
DR Pfam; PF00071; ras; 1.  
DR SMART; SM00010; small\_GTPase; 1.  
FT NON\_TER 1 56  
FT NON\_TER 56  
SQ SEQUENCE 56 AA; 6283 MW; B3DBFC0450453726 CRC64;  
Query Match 100.0%; Score 20; DB 5; Length 56;  
Best Local Similarity 100.0%; Pred. No. 5.4e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 IEGR 4  
Db 39 IEGR 42  
RESULT 5  
Q9UPG5  
ID Q9UPG5 PRELIMINARY; PRT; 58 AA.  
AC Q9UPG5;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Serine-threonine kinase (Fragment).  
GN BTAK.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BREAST;  
RA Sen S., Zhou H.;  
RT "Partial Genomic Sequence of BTAK.";  
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF011467; AAC23448.1; -.  
DR InterPro; IPR000719; Euk\_pkinase.  
DR Pfam; PF00069; pkinase; 1.  
DR ProDom; PD000001; Euk\_pkinase; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
KW Kinase.  
FT NON\_TER 1 58  
FT NON\_TER 58  
SQ SEQUENCE 58 AA; 6812 MW; 30D49CBF5F6B5F86 CRC64;  
Query Match 100.0%; Score 20; DB 4; Length 58;  
Best Local Similarity 100.0%; Pred. No. 5.6e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 IEGR 4  
Db 16 IEGR 19  
RESULT 6  
Q93U42  
ID Q93U42 PRELIMINARY; PRT; 60 AA.  
AC Q93U42;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)

DE RNA polymerase beta subunit (Fragment).  
GN RPOB.  
OS Carsonella ruddii.  
OC Bacteria; Proteobacteria; gamma subdivision; Candidatus Carsonella.  
OX NCBI\_TaxID=114186;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Thao M.L., Clark M.A., Burckhardt D.H., Moran N.A., Baumann P.;  
RT "Phylogenetic analysis of vertically transmitted psyllid endosymbionts  
RT (Candidatus Carsonella ruddii) based on atpAGD and rpoC; comparisons  
RT with 16S-23S rDNA-derived phylogeny.";  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
CC SUBSTRATES (BY SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE - N DIPHOSPHATE +  
CC [RNA](N).  
CC -!- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE  
CC ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1  
CC BETA' CHAIN (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.  
CC EMBL; AF268063; AAK55952.1; -.  
DR InterPro: IPR001572; RNA\_pol\_B.  
DR Pfam: PF00562; RNA\_pol\_B; 1.  
KW DNA-directed RNA polymerase; Transcription; Transferase.  
FT NON\_TER 1  
SQ SEQUENCE 60 AA; 6786 MW; 90BB0074478AFBFD CRC64;

Query Match 100.0%; Score 20; DB 2; Length 60;  
Best Local Similarity 100.0%; Pred. No. 5.8e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
Db 18 IEGR 21

RESULT 7  
Q93U40 PRELIMINARY; PRT; 60 AA.  
AC Q93U40;  
DT 01-DEC-2001 (TRENBLrel. 19, Created)  
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)  
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)  
DE RNA polymerase beta subunit (Fragment).  
GN RPOB.  
OS Carsonella ruddii.  
OC Bacteria; Proteobacteria; gamma subdivision; Candidatus Carsonella.  
OX NCBI\_TaxID=114186;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Thao M.L., Clark M.A., Burckhardt D.H., Moran N.A., Baumann P.;  
RT "Phylogenetic analysis of vertically transmitted psyllid endosymbionts  
RT (Candidatus Carsonella ruddii) based on atpAGD and rpoC; comparisons  
RT with 16S-23S rDNA-derived phylogeny.";  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
CC SUBSTRATES (BY SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE - N DIPHOSPHATE +  
CC [RNA](N).  
CC -!- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE  
CC ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1  
CC BETA' CHAIN (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.  
CC EMBL; AF268063; AAK55954.1; -.  
DR InterPro: IPR001572; RNA\_pol\_B.  
DR Pfam: PF00562; RNA\_pol\_B; 1.  
KW DNA-directed RNA polymerase; Transcription; Transferase.  
FT NON\_TER 1  
SQ SEQUENCE 60 AA; 6617 MW; 9A34D68E7E563350 CRC64;

Query Match 100.0%; Score 20; DB 2; Length 60;

Best Local Similarity 100.0%; Pred. No. 5.8e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
Db 18 IEGR 21

RESULT 8  
Q91GA0 PRELIMINARY; PRT; 61 AA.  
AC Q91GA0;  
DT 01-DEC-2001 (TRENBLrel. 19, Created)  
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)  
DE Hypothetical 6.7 kDa protein.  
CC columbia circovirus.  
OS Viruses; ssDNA viruses; Circoviridae; Circovirus.  
OX NCBI\_TaxID=126070;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=9030;  
RA Weston J.H.;  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=9030;  
RC MEDLINE=21378871; PubMed=11485403;  
RA Todd D., Weston J.H., Soike D., Smyth J.A.;  
RT "Genome sequence determinations and analyses of novel circoviruses  
RT from goose and pigeon.";  
RL Virology 286:354-362(2001).  
DR EMBL; AJ298229; CAC50248.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 61 AA; 6736 MW; 99F8DFF347A1A332 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 61;  
Best Local Similarity 100.0%; Pred. No. 5.9e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
Db 15 IEGR 18

RESULT 9  
Q93JW0 PRELIMINARY; PRT; 63 AA.  
AC Q93JW0;  
DT 01-DEC-2001 (TRENBLrel. 19, Created)  
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)  
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
DE 4-oxalocrotonate tautomerase.  
GN TDNL.  
OS Pseudomonas putida.  
OG Plasmid pTDNL.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=303;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Fukumori F.;  
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; D85415; BAB62059.1; -.  
DR InterPro: IPR004370; Taut.  
DR Pfam: PF01361; Tautomerase; 1.  
DR ProDom: PD019232; Tautomerase; 1.  
DR TIGRFAMs: TIGR00013; taut; 1.  
KW Plasmid.  
SQ SEQUENCE 63 AA; 6927 MW; 97EDA73A49AED63B CRC64;

Query Match 100.0%; Score 20; DB 2; Length 63;  
Best Local Similarity 100.0%; Pred. No. 6.1e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
||||  
DB 9 IEGR 12

## RESULT 10

Q9AP90 PRELIMINARY; PRT; 66 AA.  
AC Q9AP90;  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE Hypothetical 7.4 kDa protein.  
OS Uncultured bacterium F01.  
OC Bacteria; environmental samples.  
OX NCBI\_TaxID=147490;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TRANSPOSON-CLASS 7 INTEGRON;  
RX MEDLINE=21100264; PubMed=11166996;  
RA Nield B.S., Holmes A.J., Gillings M.R., Recchia G.D., Mabbutt B.C.,  
RT Nevalainen K.M.H., Stokes H.W.;  
RT "Recovery of new integron classes from environmental DNA.";  
RL FEMS Microbiol. Lett. 195:59-65(2001).  
DR EMBL: AF3141190; AAK00306.1; -  
KW Hypothetical protein.  
SQ SEQUENCE 66 AA; 7404 MW; B132766C53690D31 CRC64;

Query Match 100.0%; Score 20; DB 2; Length 66;

Best Local Similarity 100.0%; Pred. No. 6.5e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
||||  
DB 28 IEGR 31

## RESULT 11

O51760 PRELIMINARY; PRT; 66 AA.  
ID O51760  
AC O51760;  
DT 01-JUN-1998 (TREMBlrel. 06, Created)  
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
DE Hypothetical protein BB0820.  
GN BB0820.  
OS Borrelia burgdorferi (Lyme disease spirochete).  
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.  
OX NCBI\_TaxID=1439;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 35210 / B31;  
RX MEDLINE=98065943; PubMed=9403685;  
RA Fraser C.M., Castens S., Huang W.M., Sutton G.G., Clayton R.A.,  
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwin M.,  
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,  
RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,  
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,  
RA Utterback T., Watthey L., McDonald L., Artiach P., Bowman C.,  
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,  
RA Smith H.O., Venter J.C.;  
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.";  
RL Nature 390:580-586(1997).  
DR EMBL: AF001180; AAC67174.1; -  
DR TIGR: BB0820; -  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 66 AA; 7579 MW; 21352F7A53946F45 CRC64;

Query Match 100.0%; Score 20; DB 16; Length 66;

Best Local Similarity 100.0%; Pred. No. 6.5e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
||||  
DB 63 IEGR 66

## RESULT 12

O8YGJ9 PRELIMINARY; PRT; 67 AA.  
AC O8YGJ9;  
DT 01-MAR-2002 (TREMBlrel. 20, Created)  
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
DE Hypothetical protein BME11159.  
GN BME11159.  
OS Brucella melitensis.  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Brucellaceae; Brucella.  
OX NCBI\_TaxID=29459;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=16M / ATCC 23456 / BIOTYPE 1;  
RX MEDLINE=20020109; PubMed=11756688;  
RA DelVecchio V.G., Kapatral V., Redkar R.J., Patra G., Mijer C., Los T.,  
RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,  
RA Jablonki L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltsman E.,  
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,  
RA Haselkorn R., Kyrpides N., Overbeek R.;  
RT "The genome sequence of the facultative intracellular pathogen Brucella melitensis.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).  
DR EMBL: AF009555; AAL52340.1; -  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 67 AA; 7381 MW; 710993A78305185B CRC64;

Query Match 100.0%; Score 20; DB 16; Length 67;

Best Local Similarity 100.0%; Pred. No. 6.6e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
||||  
DB 37 IEGR 40

## RESULT 13

O38263 PRELIMINARY; PRT; 68 AA.  
ID O38263  
AC O38263;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE Unidentified ORF20.  
OS Lactococcus phage b167.  
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;  
OC c2-like viruses.  
OX NCBI\_TaxID=36343;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95111629; PubMed=7812447;  
RA Schouler C., Ehrlich S.D., Chopin M.C.;  
RT "Sequence and organization of the lactococcal prolate-headed b167 phage genome.";  
RL Microbiology 140:3061-3069(1994).  
DR EMBL: L33769; AAA74357.1; -  
SQ SEQUENCE 68 AA; 8153 MW; D096C8C0B41BCF13 CRC64;

Query Match 100.0%; Score 20; DB 9; Length 68;

Best Local Similarity 100.0%; Pred. No. 6.7e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
||||



Db 27 IEGR 30

## RESULT 14

Q8XRG1

ID Q8XRG1 PRELIMINARY; PRT; 70 AA.

AC Q8XRG1; 20, Created)  
 DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
 DE Probable 4-oxalocrotonate isomerase protein (EC 5.3.2.-).  
 GN RSP0893 OR RS01664.

OS Ralstonia solanacearum (Pseudomonas solanacearum).  
 OC Plasmid megaplasmid.  
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
 OC Ralstonia.

OX NCBI\_TaxID=305;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=GMT1000;

RX MEDLINE=21681879; PubMed=11823852;  
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
 RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,  
 RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,  
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schliex T.,  
 RA Siquier P., Thebault P., Whalen M., Wincker P., Levy M.,  
 RA Weissenbach J., Boucher C.A.;

RT "Genome sequence of the plant pathogen Ralstonia solanacearum";  
 RL Nature 415:497-502(2002).  
 DR EMBL; AL646081; CAD18044.1; -;  
 DR InterPro; IPR004370; Taut.

DR Pfam; PF01361; Tautomerase; 1.  
 DR ProDom; PD019232; Tautomerase; 1.  
 DR TIGRFAMs; TIGR00013; taut; 1.

KW Isomerase; Plasmid; Complete proteome.  
 SQ SEQUENCE 70 AA; 7256 MW; 2D9DED5IA9C4ABCA CRC64;

Query Match 100.0%; Score 20; DB 16; Length 70;  
 Best Local Similarity 100.0%; Pred. No. 6.9e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IEGR 4  
 |||||

Db 9 IEGR 12

## RESULT 15

Q9PDH9

ID Q9PDH9 PRELIMINARY; PRT; 71 AA.

AC Q9PDH9; 20, Created)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
 DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)  
 DE Hypothetical protein xfl1400.  
 GN Xfl1400.

OS Xylella fastidiosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
 OC Xylella.

OX NCBI\_TaxID=2371;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=9A5C;

RX MEDLINE=20365717; PubMed=10910347;  
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,  
 RA Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,  
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,  
 RA Bueno M.R.P., Camargo A.A., Carraro D.M., Carrier H.,  
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,  
 RA Facincani L.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,  
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,  
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
 RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,

RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,  
 RA Marcho M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,  
 RA Menck C.F.M., Mirsacca E.C., Miyaki C.Y., Monteliro-Vitorello C.B.,  
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
 RA Nhami A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Pesquero J.B.,  
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., de Souza A.A.,  
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,  
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,  
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;  
 RT "The genome sequence of the plant pathogen Xylella fastidiosa";  
 RL Nature 406:151-159(2000).  
 DR EMBL; AE003971; AAF84209.1; -;  
 KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 71 AA; 7647 MW; A85C21C5DBF3FA0E CRC64;

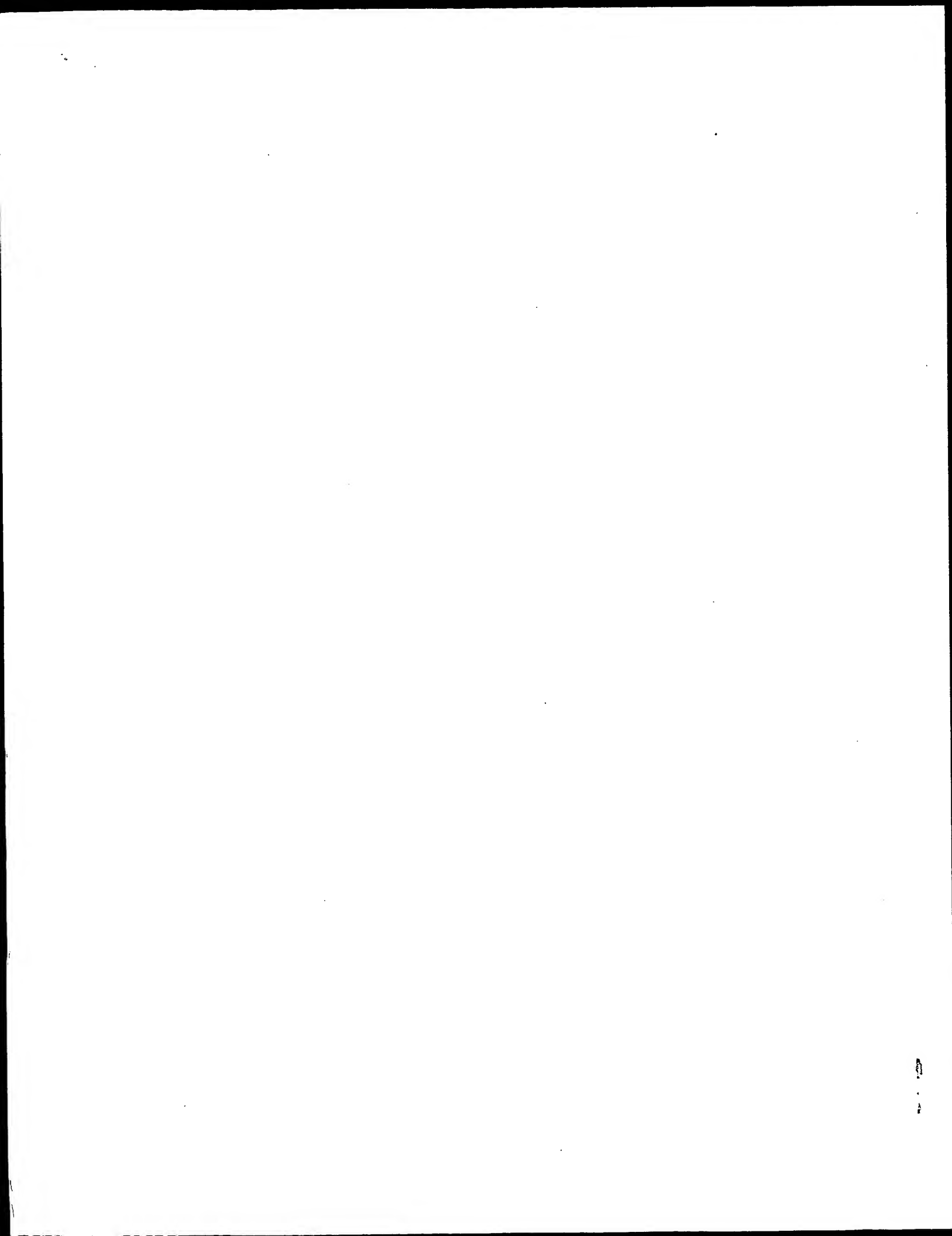
Query Match 100.0%; Score 20; DB 16; Length 71;  
 Best Local Similarity 100.0%; Pred. No. 7e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IEGR 4  
 |||||

Db 24 IEGR 27

Search completed: April 13, 2003, 03:22:41

Job time: 2.2128 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2003, 01:36:25 ; Search time 0.0764755 Seconds  
(without alignments)  
2169.392 Million cell updates/sec

Title: US-10-023-888-22

Perfect score: 20

Sequence: 1 IEGR 4

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	IF2A_RABIT	20	100.0	52	1	IF2A_RABIT	P83268 Oryctolagus
2	DMP1_PSESP	20	100.0	62	1	DMP1_PSESP	P49172 pseudomonas
3	Y889_ARCFU	20	100.0	73	1	Y889_ARCFU	O28584 archaeoglob
4	YAJ9_YEAST	20	100.0	74	1	YAJ9_YEAST	P39549 saccharomyc
5	ACP_BACHD	20	100.0	77	1	ACP_BACHD	O9ka04 bacillus ha
6	RL27_RICCN	20	100.0	86	1	RL27_RICCN	O92990 rickettsia
7	RL27_RICPR	20	100.0	86	1	RL27_RICPR	O92990 rickettsia
8	YN77_CAUCR	20	100.0	105	1	YN77_CAUCR	O9a5s0 caulobacter
9	BPA3_PSES1	20	100.0	109	1	BPA3_PSES1	O52440 pseudomonas
10	PYRE_ENTFA	20	100.0	112	1	PYRE_ENTFA	O07657 enterococcu
11	Y211_METUA	20	100.0	114	1	Y211_METUA	O60273 methanococc
12	ACYP_DROME	20	100.0	119	1	ACYP_DROME	P56544 drosophila
13	SY07_PIG	20	100.0	119	1	SY07_PIG	P43030 sus scrofa
14	YRDN_BACSU	20	100.0	129	1	YRDN_BACSU	P45032 bacillus su
15	V56_HPV49	20	100.0	138	1	V56_HPV49	P36813 human papil
16	FLAV_TREPA	20	100.0	146	1	FLAV_TREPA	O38395 treponema p
17	NORC_PARDE	20	100.0	149	1	NORC_PARDE	O51662 paracoccus
18	PYRI_PYRAE	20	100.0	151	1	PYRI_PYRAE	O8ztg2 pyrobaculum
19	HSPA_BRAJA	20	100.0	152	1	HSPA_BRAJA	P70917 bradyrhizob
20	C554_RHOSH	20	100.0	153	1	C554_RHOSH	O33142 rhodobacter
21	CHEW_CAUCR	20	100.0	155	1	CHEW_CAUCR	O87715 caulobacter
22	CHEW_RHIME	20	100.0	155	1	CHEW_RHIME	O52881 rhizobium m
23	GREM_LACIA	20	100.0	156	1	GREM_LACIA	O9cht2 lactococcus
24	YFEB_LACIA	20	100.0	156	1	YFEB_LACIA	O48660 lactococcus
25	FABL_CAEEL	20	100.0	161	1	FABL_CAEEL	O20224 caenorhabdi
26	GREM_RICPR	20	100.0	162	1	GREM_RICPR	P27640 rickettsia
27	Y22B_HAEIN	20	100.0	163	1	Y22B_HAEIN	O86223 haemophilus
28	YB38_MYCPN	20	100.0	166	1	YB38_MYCPN	P75260 mycoplasma
29	PYRE_HALN1	20	100.0	175	1	PYRE_HALN1	O9hng2 halobacteri
30	IF3_BUCAP	20	100.0	178	1	IF3_BUCAP	P46243 buchnera ap
31	IF3_BUCAI	20	100.0	179	1	IF3_BUCAI	P37226 buchnera ap
32	IF3_PROVO	20	100.0	179	1	IF3_PROVO	P33319 proteus vul
33	SPC4_HUMAN	20	100.0	179	1	SPC4_HUMAN	P21337 homo sapien

34	20	100.0	179	1	SPC4_MOUSE	Q9r0p6 mus musculu
35	20	100.0	179	1	SPC4_RAT	P42667 rattus norv
36	20	100.0	179	1	SSB_HELPY	O25841 helicobacte
37	20	100.0	180	1	IF3_ECOLI	P02999 escherichia
38	20	100.0	180	1	IF3_KLEPN	P33318 klebsiella
39	20	100.0	180	1	IF3_SALTI	O82613 salmonella
40	20	100.0	180	1	IF3_SALTI	P33321 salmonella
41	20	100.0	181	1	KCY_PYRAB	O9uzj6 pyrococcus
42	20	100.0	181	1	SSB_HELPJ	O9zjy2 helicobacte
43	20	100.0	182	1	Y4YS_RHISN	P55727 rhizobium s
44	20	100.0	183	1	HGXR_TRIFO	P51900 tritrichomo
45	20	100.0	183	1	IF3_SERNA	P33320 serratia ma

## ALIGNMENTS

RESULT 1  
IF2A\_RABIT  
ID IF2A\_RABIT STANDARD; PRT; 52 AA.  
AC P83268;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Eukaryotic translation initiation factor 2 subunit 1 (Eukaryotic  
DE translation initiation factor 2 alpha subunit) (eif-2-alpha) (EIF-  
DE 2alpha) (EIF-2A) (Fragment).  
DE  
GN EIF2S1 OR EIF2A.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_TaxID=9986;  
RN [1]  
RP SEQUENCE, AND PHOSPHORYLATION OF SER-48.  
RC TISSUP-Reticulocytes;  
RX MEDLINE=8630412; PubMed=3745199;  
RA Wettenhall R.E.H., Kudlicki W., Kramer G., Hardesty B.:  
RT "The NH2-terminal sequence of the alpha and gamma subunits of  
RT eukaryotic initiation factor 2 and the phosphorylation site for the  
RT heme-regulated eif-2 alpha kinase.";  
RL J. Biol. Chem. 261:12444-12447(1986).  
RL [2]  
CC SEQUENCE OF 1-23.  
CC TISSUP-Reticulocytes;  
CC MEDLINE=87240215; PubMed=3592677;  
CC Schaefer M.P., Fairwell T., Parker D.S., Knight M., Anderson W.F.,  
CC Safer B.:  
CC "The purification and characterization of subunits alpha, beta, and  
CC gamma from the rabbit reticulocyte eukaryotic initiation factor 2.";  
CC Arch. Biochem. Biophys. 255:337-346(1987).  
CC -1- FUNCTION: Functions in the early steps of protein synthesis by  
CC forming a ternary complex with GTP and initiator tRNA. This  
CC complex binds to a 40s ribosomal subunit, followed by mRNA binding  
CC to form a 43S preinitiation complex. Junction of the 60S ribosomal  
CC subunit to form the 80S initiation complex is preceded by  
CC hydrolysis of the GTP bound to eIF-2 and release of an eIF-2-GDP  
CC binary complex. In order for eIF-2 to recycle and catalyze another  
CC round of initiation, the GDP bound to eIF-2 must exchange with GTP  
CC by way of a reaction catalyzed by eIF-2B.  
CC -1- SUBUNIT: Heterotrimer composed of an alpha, a beta, and a gamma  
CC chain.  
CC -1- PTM: Phosphorylation at least 4 kinases: EIF2AK3/PERK, GCN2, HRI and  
CC PKR. Phosphorylation stabilizes the eIF-2/GDP/eIF-2B complex and  
CC prevents GDP/GTP exchange reaction, thus impairing the recycling  
CC of eIF-2 between successive rounds of initiation and leading to  
CC global inhibition of translation (By similarity).  
CC -1- SIMILARITY: BELONGS TO THE EIF-2-ALPHA FAMILY.  
CC -1- SIMILARITY: CONTAINS 1 SI MOTIF DOMAIN.  
CC InterPro: IPR003029; SI.  
CC Pfam: PF00575; SI; 1.  
CC PROSITE: PS0126; SI; 1.  
CC Initiation factor; Protein biosynthesis;  
CC Translation regulation; RNA-binding; Phosphorylation.

```
FT DOMAIN 16 >52 S1 MOTIF.
FT MOD_RES 48 48 PHOSPHORYLATION (BY HRI).
FT MOD_RES 51 51 PHOSPHORYLATION (BY EIF2AK3, GCN2, HRI
FT AND PKR) (BY SIMILARITY).
FT UNSURE 10 10
FT UNSURE 44 44
FT UNSURE 52 52
FT CONFLICT 10 10 H -> R (IN REF. 2).
FT CONFLICT 16 16 E -> Q (IN REF. 1).
FT NON_TER 52 52
SQ SEQUENCE 52 AA; 5974 MW; 52E63D8DCEA6B804 CRC64;

Query Match 100.0%; Score 20; DB 1; Length 52;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4
DB 41 IEGR 44

RESULT 2
DMPI_PSESP STANDARD; PRT; 62 AA.
AC P49172;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 4-oxalocrotonate tautomerase (EC 5.3.2.-) (4-OT).
GN DMPI.
OS Pseudomonas sp. (strain CF600).
OG Plasmid pVil50.
OC Bacteria; Proteobacteria.
OX NCBI_TaxID=306;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92121108; PubMed=1732207;
RA Shingler V., Marklund U., Powlowski J.;
RT "Nucleotide sequence and functional analysis of the complete
RT phenol/3,4-dimethylphenol catabolic pathway of Pseudomonas sp.
RT strain CF600."
RL J. Bacteriol. 174:711-724 (1992).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX MEDLINE=96146412; PubMed=8547259;
RA Subramanya H.S., Roper D.I., Dauter Z., Dodson E.J., Davies G.J.,
RA Wilson K.S., Wigley D.B.;
RT "Enzymatic ketonization of 2-hydroxymuconate: specificity and
RT mechanism investigated by the crystal structures of two isomerases."
RL Biochemistry 35:792-802 (1996).
CC -!- FUNCTION: CATALYZES THE KETONIZATION OF 2-HYDROXYMUCONATE
CC STEREOSELECTIVELY TO YIELD 2-OXO-3-HEXENEDIOATE.
CC -!- PATHWAY: META-CLEAVAGE PATHWAY FOR THE DEGRADATION OF TOLUENE,
CC M-XYLENE AND P-XYLENE.
CC -!- SUBUNIT: HOMOHexamER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X60835; CAA43229.1; -.
CC PDB; 1OTF; 03-APR-96.
CC InterPro; IPR004370; Taut.
CC Pfam; PF01361; Tautomerase; 1.
CC ProDom; PD019232; Tautomerase; 1.
CC TIGRFAMs; TIGR00013; taut; 1.
CC Isomerase; Plasmid; Aromatic hydrocarbons catabolism; 3D-structure.
FT INIT_MET 0 0 BY SIMILARITY.
FT ACT_SITE 1 1 CATALYTIC BASE (BY SIMILARITY).
```

```
SQ SEQUENCE 62 AA; 6974 MW; CEDAB0F92D1E1E0F CRC64;

Query Match 100.0%; Score 20; DB 1; Length 62;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4
DB 8 IEGR 11

RESULT 3
YG89_ARCFU STANDARD; PRT; 73 AA.
AC O28584;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AFI689.
GN AFI689.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Cocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370 (1997).
CC -!- SIMILARITY: TO M.JANNASCHII MJ122 AND AF1997.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF000987; AAB89570.1; -.
CC TIGR; AFL689; -.
CC InterPro; IPR003847; DUF217.
CC Pfam; PF02697; DUF217; 1.
CC KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 73 AA; 8782 MW; D059A3E6B02452C3 CRC64;

Query Match 100.0%; Score 20; DB 1; Length 73;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4
DB 32 IEGR 35

RESULT 4
YAJ9_YEAST STANDARD; PRT; 74 AA.
ID YAJ9_YEAST
AC P39549;
DT 01-FEB-1995 (Rel. 31, Created)
```

DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical 8.9 kDa protein in CDC15-YAT1 intergenic region.  
GN YAR029W OR FUN57.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=5288c / AB972;  
RA Bussey H., Keng T., Storms R.K., Vo D.T., Clark M.W., Fortin N.,  
RA Barton A.B., Kaback D.B., Clark M.W.;  
RL Submitted (FEB-1994) to the EMBL/GenBank/DDJB databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=5288c / AB972;  
RX MEDLINE=95249563; PubMed=7731988;  
RA Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,  
RA Hall J., Ouellette B.F.F., Keng T., Barton A.B., Su Y., Davies C.K.,  
RA Storms R.K.;  
RT "The nucleotide sequence of chromosome I from Saccharomyces  
RT cerevisiae.";  
RL Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).  
CC -1- SIMILARITY: BELONGS TO THE DUF/COS FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: L28920; AAC09491.1; -  
CC SGD: S0000077; YAR029W.  
CC InterPro: IPR001142; DUF.  
CC Pfam: PF00674; DUF; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 74 AA; 8910 MW; 11CDFB4F582BBF21 CRC64;  
Query Match 100.0%; Score 20; DB 1; Length 74;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 IEGR 4  
DB 31 IEGR 34  
RESULT 5  
ACP\_BACHD  
ID ACP\_BACHD STANDARD; PRT; 77 AA.  
AC Q9KA04;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Acyl carrier protein (ACP).  
GN ACPP OR ACPA OR BH2490.  
OS Bacillus halodurans.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=86665;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C-125 / JCM 9153;  
RX MEDLINE=20512582; PubMed=11058132;  
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,  
RA Horikoshi K.;  
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
RT halodurans and genomic sequence comparison with Bacillus subtilis.";  
RL Nucleic Acids Res. 28:4317-4331(2000).  
CC -1- FUNCTION: Carrier of the growing fatty acid chain in fatty

CC acid biosynthesis (By similarity).  
CC -1- PATHWAY: De novo fatty acid biosynthesis.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -1- FM: 4'-phosphopantetheine is transferred from CoA to a specific  
CC serine of apo-ACP by acps. This modification is essential for  
CC activity because fatty acids are bound in thioester linkage to the  
CC sulfhydryl of the prosthetic group (By similarity).  
CC -1- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: AP001515; BAB06209.1; -  
CC HSP: P80643; LHY8.  
CC InterPro: IPR003231; Acyl\_carrier.  
CC InterPro: IPR003880; Pantne\_attach.  
CC Pfam: PF00550; pp-binding; 1.  
CC ProDom: PD000887; Acyl\_carrier; 1.  
CC TIGRFAMs: TIGR00517; acyl\_carrier; 1.  
CC PROSITE: PS00075; ACP\_DOMAIN; 1.  
CC PROSITE: PS00012; PHOSPHOPANTHEINE; 1.  
KW Lipid synthesis; Fatty acid biosynthesis; Phosphopantetheine;  
KW Complete proteome.  
SQ SEQUENCE 77 AA; 8538 MW; 3575CDFAM5BA15CA CRC64;  
Query Match 100.0%; Score 20; DB 1; Length 77;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 IEGR 4  
DB 73 IEGR 76  
RESULT 6  
RL27\_RICCN  
ID RL27\_RICCN STANDARD; PRT; 86 AA.  
AC Q92GGO;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE 50S ribosomal protein L27.  
GN RPLA OR RCL163.  
OS Rickettsia conorii.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Rickettsiaceae; Rickettsieae; Rickettsia.  
OX NCBI\_TaxID=781;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Malish 7;  
RX MEDLINE=21442074; PubMed=11557893;  
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,  
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,  
RA Raoult D.;  
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";  
RL Science 293:2093-2098(2001).  
CC -1- SIMILARITY: BELONGS TO THE L27P FAMILY OF RIBOSOMAL PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: AE008664; AAL03701.1; -

```

DR InterPro: IPR001684; Ribosomal_L27.
DR Pfam: PF01016; Ribosomal_L27; 1.
DR ProDom: PD003114; Ribosomal_L27; 1.
DR TIGRFAMs: TIGR00062; L27; 1.
DR PROSITE: PS00831; RIBOSOMAL_L27; 1.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 86 AA; 9293 MW; 97ED818F999658C4 CRC64;

Query Match 100.0%; Score 20; DB 1; Length 86;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEGR 4
Db 63 IEGR 66

RESULT 7
RL27_RICPR STANDARD; PRT; 86 AA.
AC Q9ZCIG;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 50S ribosomal protein L27.
GN RPMA OR RP752.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Madrid E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sichteritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria.";
RL Nature 396:133-140(1998).
CC -!- SIMILARITY: BELONGS TO THE L27P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AJ235273; CAA15180.1; -
DR InterPro: IPR001684; Ribosomal_L27.
DR Pfam: PF01016; Ribosomal_L27; 1.
DR PRINTS: PR00063; RIBOSOMAL_L27.
DR ProDom: PD003114; Ribosomal_L27; 1.
DR TIGRFAMs: TIGR00062; L27; 1.
DR PROSITE: PS00831; RIBOSOMAL_L27; 1.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 86 AA; 9307 MW; 15EF50BD91BE79D8 CRC64;

Query Match 100.0%; Score 20; DB 1; Length 86;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEGR 4
Db 63 IEGR 66

RESULT 8
YN77_CAUCR
ID YN77_CAUCR STANDARD; PRT; 105 AA.
AC Q9A5S0;

DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Hypothetical protein CC2377.
GN CC2377.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke M.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
CC -!- SIMILARITY: BELONGS TO THE UPF0145 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AE005907; AAK24348.1; -
DR TIGR; CC2377; -
DR InterPro: IPR002765; DUF74.
DR Pfam: PF01906; DUF74; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 105 AA; 11269 MW; 74EC7E6FEF73701A CRC64;

Query Match 100.0%; Score 20; DB 1; Length 105;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEGR 4
Db 9 IEGR 12

RESULT 9
BPA3_PSES1 STANDARD; PRT; 109 AA.
AC Q52440;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Ferredoxin subunit of biphenyl dioxygenase.
GN BPA3.
OS Pseudomonas sp. (strain KKS102).
OC Bacteria; Proteobacteria.
OX NCBI_TaxID=307;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94324977; PubMed=8048958;
RA Fukuda M., Yasukouchi Y., Kikuchi Y., Nagata Y., Kimbara K.,
RA Horiuchi H., Takagi M., Yano K.;
RT "Identification of the bpha and bphB genes of Pseudomonas sp. strains
RT KKS102 involved in degradation of biphenyl and polychlorinated
RT biphenyls.";
RL Biochem. Biophys. Res. Commun. 202:850-856(1994).
CC -!- FUNCTION: THIS PROTEIN SEEMS TO BE A 2FE-2S FERREDOXIN.
CC -!- SUBUNIT: THIS DIOXYGENASE SYSTEM CONSISTS OF FOUR PROTEINS: THE
CC TWO SUBUNITS OF THE HYDROXYLASE COMPONENT (BPHA1 AND BPHA2), A
CC FERREDOXIN (BPHA3) AND A FERREDOXIN REDUCTASE (BPHA4).

```



CC -1- SIMILARITY: BELONGS TO THE BACTERIAL RING-HYDROXYLATING  
 CC DIOXYGENASE FERREDOXIN COMPONENT FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: D17319; BAA04139.1; -  
 DR HSSP: P37332; IFOT.  
 DR InterPro: IPR001281; Rieske.  
 DR Pfam: PF00355; Rieske; 1.  
 KW Aromatic hydrocarbons catabolism; Electron transport; Iron-sulfur.  
 FT METAL 43 43 IRON-SULFUR (2FE-2S) (POTENTIAL).  
 FT METAL 45 45 IRON-SULFUR (2FE-2S) (POTENTIAL).  
 FT METAL 63 63 IRON-SULFUR (2FE-2S) (POTENTIAL).  
 FT METAL 66 66 IRON-SULFUR (2FE-2S) (POTENTIAL).  
 SQ SEQUENCE 109 AA; 11912 MW; 0D3BB8C60C01751 CRC64;  
 Query Match 100.0%; Score 20; DB 1; Length 109;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IEGR 4  
 Db 93 IEGR 96  
 RESULT 10  
 ID PYRE\_ENTFA STANDARD; PRT; 112 AA.  
 AC 007657;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Orotate phosphoribosyltransferase (EC 2.4.2.10) (OPRT) (OPRTase)  
 DE (Fragment).  
 GN PYRE.  
 OS Enterococcus faecalis (Streptococcus faecalis).  
 OG Plasmid pKV48.  
 OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.  
 OX NCBI\_TaxID=1351;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 47077 / OGIRF;  
 RX MEDLINE=96074317; PubMed=7592480;  
 RA Li X., Weinstock G.M., Murray B.E.;  
 RT "Generation of auxotrophic mutants of Enterococcus faecalis.";  
 RL J. Bacteriol. 177:6866-6873(1995).  
 CC -1- CATALYTIC ACTIVITY: Orotidine 5'-phosphate + diphosphate = orotate  
 CC + 5-phospho-alpha-D-ribose 1-diphosphate.  
 CC -1- PATHWAY: Pyrimidine biosynthesis; fifth step.  
 CC -1- SIMILARITY: BELONGS TO THE PURINE/PYRIMIDINE  
 CC PHOSPHORIBOSYLTRANSFERASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U24682; AAB61216.1; -  
 DR HSSP: P08870; IOPR.  
 DR InterPro: IPR002375; Pr/pv\_rp.transf.  
 DR InterPro: IPR000836; PRtransferase.  
 DR Pfam: PF00156; Ribosyltran; 1.  
 DR PROSITE: PS00103; PUR\_PIR\_PR\_TRANSFER; 1.  
 KW Pyrimidine biosynthesis; Transferase; Glycosyltransferase; Plasmid.

FT NON\_TER 1 1  
 FT ACT\_SITE 75 75 BY SIMILARITY.  
 FT NON\_TER 112 112  
 SQ SEQUENCE 112 AA; 11910 MW; 5B88F43A8D3AF698 CRC64;  
 Query Match 100.0%; Score 20; DB 1; Length 112;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IEGR 4  
 Db 83 IEGR 86  
 RESULT 11  
 ID YZ11\_METJA STANDARD; PRT; 114 AA.  
 AC 060273;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein MJEC11.  
 GN MJEC11.  
 OS Methanococcus jannaschii.  
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
 OC Methanocaldococcaceae; Methanocaldococcus.  
 OX NCBI\_TaxID=2190;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=96337999; PubMed=8688087;  
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kervage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
 RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 RT jannaschii.";  
 RL Science 273:1058-1073(1996).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: L77118; AAC37084.1; -  
 DR TIGR: MJEC11; -  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 114 AA; 12826 MW; 99A168C6180676B CRC64;  
 Query Match 100.0%; Score 20; DB 1; Length 114;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IEGR 4  
 Db 77 IEGR 80  
 RESULT 12  
 ID ACYP\_DROME STANDARD; PRT; 119 AA.  
 AC P56544; Q9V3K1;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Acylphosphatase (EC 3.6.1.7) (Acylphosphate phosphohydrolase).

GN ACYP OR ACP OR ACPPRO OR CG16870.  
 OS Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 CC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 CC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98416044; PubMed=9744795;  
 RA Pieri A., Magherini F., Liquori G., Raugeri G., Taddei N.,  
 RA Bozzetti M.P., Cecchi C., Ramponi G.;  
 RT "Drosophila melanogaster acylphosphatase: a common ancestor for  
 RT acylphosphatase isoenzymes of vertebrate species.";  
 RL FEBS Lett. 433:205-210(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RX MEDLINE=99403001; PubMed=10471707;  
 RA Ashburner M., Misra S., Rote J., Lewis S.E., Blazej R., Davis T.,  
 RA Doyle C., Galle R., George R., Harris N., Hartzell G., Harvey D.,  
 RA Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,  
 RA Palazzolo M., Reese M.G., Spradling A., Tsang G., Wan K., Whitelaw K.,  
 RA Celniker S., Rubin G.M.;  
 RT "An exploration of the sequence of a 2.9-Mb region of the genome of  
 RT Drosophila melanogaster: the Adh region.";  
 RL Genetics 153:179-219(1999).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burtis K.C., Busam A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Dou L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hoston D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 CC -!- CATALYTIC ACTIVITY: An acyl phosphate + H(2)O = a fatty acid anion  
 CC + phosphate.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
 CC -!- MISCELLANEOUS: OPTIMUM PH IS 5.3-6.3.  
 CC -!- SIMILARITY: BELONGS TO THE ACYLPHOSPHATASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AJ243543; CAB48386.1; -;  
 CC EMBL: AE003408; AAF44835.1; -;  
 CC EMBL: AE003641; AAF53355.1; -;  
 CC HSSP: P00818; IAFS.  
 CC FlyBase; FBgn0025115; Acyp.  
 CC InterPro; IPR001792; Acylphosphatase.  
 CC Pfam; PF00708; Acylphosphatase; 1.  
 CC PRINTS; PR00112; ACYLPHPTASE.  
 CC ProDom; PD001884; Acylphosphatase; 1.  
 CC PROSITE; PS00150; ACYLPHOSPHATASE\_1; 1.  
 CC PROSITE; PS00151; ACYLPHOSPHATASE\_2; 1.  
 CC Acetylation; Hydrolase.  
 CC INIT\_MET 0 0 BY SIMILARITY.  
 CC MOD\_RES 1 1 ACETYLATION (BY SIMILARITY).  
 CC ACT\_SITE 22 22 POTENTIAL.  
 CC ACT\_SITE 40 40 POTENTIAL.  
 CC FT ACT\_SITE 119 AA; 13566 MW; 803197175FA37795 CRC64;  
 CC SQ SEQUENCE  
 CC Query Match 100.0%; Score 20; DB 1; Length 119;  
 CC Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
 CC Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 1 IEGR 4  
 CC Db 50 IEGR 53  
 CC  
 CC RESULT 13  
 CC SZ07\_PIG STANDARD; PRT; 119 AA.  
 CC ID SZ07\_PIG  
 CC AC P43030;  
 CC DT 01-NOV-1995 (Rel. 32, Created)  
 CC DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 CC DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 CC DE Platelet basic protein precursor (PBP) (Small inducible cytokine B7)  
 CC (CXCL7).  
 CC GN SCYB7 OR PPBP.  
 CC OS Sus scrofa (Pig).  
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 CC OX NCBI\_TaxID=9823;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A., AND SEQUENCE OF 40-119.  
 CC RC TISSUE=Platelet;  
 CC RX MEDLINE=94229068; PubMed=7513641;  
 CC RA Power C.A., Proudfoot A.E.I., Magenat E., Bacon K.B., Wells T.N.C.;  
 CC RT "Molecular cloning and characterisation of a neutrophil chemotactic  
 CC protein from porcine platelets.";  
 CC RL Eur. J. Biochem. 221:713-719(1994).  
 CC -!- FUNCTION: CHEMOATTRACTANT FACTOR FOR NEUTROPHILS.  
 CC -!- MASS SPECTROMETRY: MW=8597.5; METHOD=Electrospray; RANGE=40-119.  
 CC -!- SIMILARITY: BELONGS TO THE INTERCRINE ALPHA FAMILY (SMALL CYTOKINE  
 CC C-X-C) (CHEMOKINE CXC).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X77935; CAA54907.1; -;  
 CC HSSP: P02775; INAP.

DR InterPro; IPR001089; CXC\_chemokine.smll.  
 DR InterPro; IPR001811; Chemokine\_IL8.  
 DR Pfam; PF00048; IL8; 1.  
 DR PRINTS; PR00437; SMALLCYTKXC.  
 DR SMART; SM00199; SCY; 1.  
 DR PROSITE; PS00471; SMALL\_CYTOKINES\_CXC; 1.  
 KW Cytokine; Growth factor; Chemotaxis; Mitogen; Platelet; Signal.  
 FT SIGNAL 1 33 PROBABLE.  
 FT PROPEP 34 39  
 FT CHAIN 40 119 PLATELET BASIC PROTEIN.  
 FT DISULFID 54 80 BY SIMILARITY.  
 FT DISULFID 56 96 BY SIMILARITY.  
 SQ SEQUENCE 119 AA; 12615 MW; 79E1D409CDD06B32 CRC64;

Query Match 100.0%; Score 20; DB 1; Length 119;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IEGR 4  
 DB 43 IEGR 46

RESULT 14  
 YRDN\_BACSU STANDARD; PRT; 129 AA.  
 AC P94502; O08187;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein yrdn.  
 GN YRDN.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=97175526; PubMed=9023181;  
 RA Belitsky B.R., Sonenshein A.L.;  
 RT "Altered transcription activation specificity of a mutant form of  
 Bacillus subtilis GltR, a LysR family member.";  
 RL J. Bacteriol. 179:1035-1043(1997).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=97453479; PubMed=9308178;  
 RA Sorokin A., Bolotin A., Purnelle B., Hilbert H., Lauber J.,  
 R Duesterhoeft A., Ehrlich S.D.;  
 RT "Sequence of the Bacillus subtilis genome region in the vicinity of  
 the lev operon reveals two new extracytoplasmic function RNA  
 polymerase sigma factors SigV and SigZ.";  
 RL Microbiology 143:2939-2943(1997).  
 [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Conterton I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,  
 RA Entian K.D., Errington J., Fabret C., Ferrali E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Ghim S.Y., Glaser P., Coffeau A., Golightly E.J., Grandi G.,  
 RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidine A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,

RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivoita C., Rocha E., Roche B., Rose M., Sadaie Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzensegger T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,  
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
 subtilis.";  
 RT Nature 390:249-256(1997).  
 RL Nature 390:249-256(1997).  
 CC -1- FUNCTION: PUTATIVE TARGET OF GLTR.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U79494; AAB47964.1; -;  
 DR EMBL; U79494; AAB47964.1; -;  
 DR EMBL; U93876; AAB80906.1; -;  
 DR EMBL; U93876; AAB80906.1; -;  
 DR Subtilist; BG12287; yrdn.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 129 AA; 14682 MW; 54DDEC8184E7461F CRC64;

Query Match 100.0%; Score 20; DB 1; Length 129;  
 Best Local Similarity 100.0%; Pred. No. 2e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IEGR 4  
 DB 9 IEGR 12

RESULT 15  
 VE6\_HPV49 STANDARD; PRT; 138 AA.  
 AC P36813;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE E6 protein.  
 GN E6.  
 OS Human papillomavirus type 49.  
 OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;  
 OC Papillomavirus.  
 OX NCBI\_TaxID=10616;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94265501; PubMed=8205838;  
 RA Delius H., Hofmann B.;  
 RT "Primer-directed sequencing of human papillomavirus types.";  
 RL Curr. Top. Microbiol. Immunol. 186:13-31(1994).  
 CC -1- FUNCTION: EXHIBIT A STRONG, BUT NON SPECIFIC AFFINITY FOR DOUBLE  
 CC STRANDED DNA (IN VITRO).  
 CC -1- SUBCELLULAR LOCATION: Nuclear matrix-associated.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X74480; CAA52579.1; -;

DR PIR; S36567; S36567.  
 DR InterPro; IPR001334; E6.  
 DR Pfam; PF00518; E6; 1.  
 KW Early protein; DNA-binding; Nuclear protein; Zinc-finger.  
 FT ZN\_FING 25 61  
 FT ZN\_FING 98 134  
 SQ SEQUENCE 138 AA; 16202 MW; 69AD429D88C08ADB CRC64;

Query Match 100.0%; Score 20; DB 1; Length 138;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IEGR 4  
 ||||  
 Db 83 IEGR 86

Search completed: April 13, 2003, 03:21:34  
 Job time : 1.07648 secs



XX WPI; 1984-038201/07.  
 XX Determn. of enzyme, inhibitor, activator and zymogen activity -  
 PT by reacting the enzyme with hydrolysable naphthalene derivs.  
 PT which is then allowed to form pigment with fast red-ITR-salt  
 XX  
 XX Example; Page 21; 42pp; German.  
 XX  
 CC The claimed method for the determ. of enzyme, inhibitor, activator  
 CC and zymogen activity involves reacting the enzyme with hydrolysable  
 CC naphthalene derivs. which is then allowed to form pigment with fast  
 CC red-ITR-salt. The process can be used in quality control of enzyme  
 CC preps., in clinical investigations and in the diagnosis of various  
 CC illnesses which depend on the enzyme content of blood or urine.  
 CC (Updated on 03-OCT-2002 to add missing OS field.)  
 XX  
 XX SQ Sequence 4 AA;  
 Query Match 100.0%; Score 20; DB 5; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IEGR 4  
 Db 1 IEGR 4  
 RESULT 2  
 AAP50015  
 ID AAP50015 standard; Protein; 4 AA.  
 XX  
 AC AAP50015;  
 XX  
 DT 16-AUG-2002 (updated)  
 DT 09-SEP-1991 (first entry)  
 XX  
 DE Sequence of a cleavage site which is specifically cleaved by  
 DE blood coagulation Factor Xa.  
 XX  
 KW Fusion protein; cleavage site; proteolytic cleavage.  
 XX  
 OS Homo sapiens.  
 OS Bos taurus.  
 OS  
 PN EPI61937-A.  
 XX  
 PD 21-OCT-1985.  
 XX  
 PF 15-MAY-1985; 85EP-0303414.  
 XX  
 PR 16-MAY-1984; 84GB-0012517.  
 PR 15-MAY-1985; 85GB-0012333.  
 XX  
 PA (CELL-) CELTECH LTD.  
 XX  
 PI Nagai K, Thogersen HC;  
 XX  
 DR WPI; 1985-291163/47.  
 DR N-PSDB; AAN50051, AAN50052.  
 XX  
 XX DNA sequence coding for cleavage site - which is specifically  
 PT cleaved by blood coagulation factor XA  
 XX  
 XX Claim 4; Page 23; 29pp; English.  
 XX  
 PS Vectors comprising the SQs in AAN50051 and AAN50052 are also claimed, as  
 CC is a process for the prodn. of a protein or peptide prod. in native  
 CC form which uses the vectors to produce the protein or peptide as a  
 CC fusion protein which is then cleaved with Factor Xa.  
 CC (Updated on 16-AUG-2002 to add missing OS field.)  
 XX  
 XX SQ Sequence 4 AA;

Query Match 100.0%; Score 20; DB 6; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IEGR 4  
 Db 1 IEGR 4  
 RESULT 3  
 AAR13804  
 ID AAR13804 standard; Protein; 4 AA.  
 XX  
 AC AAR13804;  
 XX  
 DT 07-NOV-1991 (firs entry)  
 DE Factor Xa substrate peptide (2).  
 DE  
 KW Assay; factor Xa; substrate; affinity.  
 XX  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Modified-site 1 /note= "methoxymalonyl-D-isoleucine"  
 FT Modified-site 2 /note= "L-Glu(OMe)"  
 FT Modified-site 4 /note= "L-Arg-p-nitroanilino"  
 FT  
 FT  
 XX WO9112338-A.  
 XX  
 PD 22-AUG-1991.  
 XX  
 PF 31-DEC-1990; 90WO-FR00974.  
 XX  
 PR 19-FEB-1990; 90FR-0001965.  
 XX  
 PA (SERB-) SERBIO.  
 XX  
 PI Quentin G, Martinoli JL;  
 XX  
 DR WPI; 1991-267149/36.  
 XX  
 PT New labelled tri: or tetra:peptide derivs. - substrates for  
 PT factor Xa assay, with better affinity, selectivity or water  
 PT solubility  
 XX  
 PS Claim 8 (j); Page 40; 46pp; French.  
 XX  
 CC The peptides represented in AAR13803-12 are examples of a generic  
 CC formula. They are substrates for assaying factor Xa (an enzyme  
 CC involved in haemostasis). Compared with known substrates they have  
 CC better affinity, selectivity and/or water solubility.  
 XX  
 XX Sequence 4 AA;  
 Query Match 100.0%; Score 20; DB 12; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IEGR 4  
 Db 1 IEGR 4  
 RESULT 4  
 AAR12870  
 ID AAR12870 standard; Protein; 4 AA.  
 XX  
 AC AAR12870;



XX 24-SEP-1991 (first entry)  
 XX Chromogenic substrate S-2860.  
 DE Bacterial endotoxin; Limulus amoebocyte clotting enzyme.  
 XX Synthetic.  
 XX Key Location/Qualifier  
 FT Modified-site 3  
 FT /label= glycolic acid  
 XX W09109052-A.  
 XX 27-JUN-1991.  
 XX 03-DEC-1990; 90WO-SE00797.  
 XX 12-DEC-1989; 89SE-0004188.  
 XX (KABI ) KABIVITRUM AB.  
 PA Arielly S;  
 PI WPI; 1991-208087/28.  
 DR New peptide derivs. - used as substrates for quantitative  
 PT determ. of bacterial endotoxins in fluids, foods or  
 PT pharmaceuticals.  
 XX Claim 3; Page 15; 21pp; English.  
 XX The N-terminal has an alpha-Ac gp., and the C-terminal has a  
 CC 4-nitroaniline (pNA) gp. The peptide is a specific example of the  
 CC generic formula: R1-A1-A2-A3-A4-R2  
 CC R1-H or a protective gp;  
 CC A1=H, Ile, Leu, or Val;  
 CC A2=Glu, Asp, Ser, or Thr;  
 CC A3=Gly or glycolic acid (Glyc);  
 CC A4=Arg or Lys; and  
 CC R2=4-nitro aniline  
 CC with the proviso that A3-Gly when A4=Lys and A3-Glyc when A4=Arg.  
 CC The peptides are used as substrates for the quantitative determ.  
 CC of bacterial endotoxins e.g. in physiological fluids, food or  
 CC pharmaceuticals. They have higher activity than the known substrate  
 CC S-2423 which has Gly-Arg as the C-terminal sequence. They are  
 CC prep'd. by solid phase synthesis.  
 CC See also AAR12867-R12869.  
 XX Sequence 4 AA;  
 SQ Query Match 100.0%; Score 20; DB 12; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IEGR 4  
 DB 1 IEGR 4  
 RESULT 5  
 AAR27101  
 ID AAR27101 standard; peptide; 4 AA.  
 XX AAR27101;  
 AC AAR27101;  
 XX 03-FEB-1993 (first entry)  
 DT Sequence of tetrapeptide recognised by factor Xa.  
 DE Factor Xa; enzymatic hydrolysis; soluble precursor.  
 XX

XX Synthetic.  
 OS EP505921-A.  
 PN 30-SEP-1992.  
 XX 19-MAR-1992; 92EP-0104753.  
 XX 28-MAR-1991; 91IT-0000861.  
 PR (ENIE ) ENIRICRCH SPA.  
 XX Galli G, Grandi G;  
 PI WPI; 1992-325240/40.  
 DR Mature polypeptide prodn. - by cleavage of soluble precursor using  
 XX immobilised factor Xa (Eng)  
 PT Claim 1; Page 5; 9pp; English.  
 PS The inventors claims a method for the preparation of a mature  
 XX polypeptide which involves the enzymatic cleavage of a soluble  
 CC precursor of the polypeptide using factor Xa. In particular the  
 CC method is used for the preparation of mature human growth hormone.  
 CC Eg, Factor Xa is immobilised on a solid insoluble support which is  
 CC brought into contact with a soluble precursor of the polypeptide,  
 CC hGH, which has the formula AAR27102-R27101-hGH. AAR27102 confers  
 CC solubility on hGH; and AAR27101 is a tetrapeptide recognised by factor  
 CC Xa.  
 XX Sequence 4 AA;  
 SQ Query Match 100.0%; Score 20; DB 13; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IEGR 4  
 DB 1 IEGR 4  
 RESULT 6  
 AAR32109  
 ID AAR32109 standard; Protein; 4 AA.  
 XX AAR32109;  
 AC AAR32109;  
 XX 11-JUL-1993 (first entry)  
 DT Sequence of blood coagulation factor Xa recognition site.  
 XX Factor Xa linker; fusion gene; blood coagulation factor.  
 XX Synthetic.  
 OS EP532043-A.  
 PN 17-MAR-1993.  
 XX 11-SEP-1992; 92EP-0115607.  
 PF 13-SEP-1991; 91JP-0234430.  
 XX (HITA ) HITACHI LTD.  
 PA Harada Y, Sakamoto T, Senda T, Takamoto K;  
 PI WPI; 1993-087069/11.  
 DR Factor Xa linker DNA - used to construct fusion genes which can  
 XX be expressed in large quantities in E coli for prodn.of foreign  
 PT

PT proteins

PS Disclosure; column 1; 12pp; English.

CC The DNA in AAQ38417 and AAQ38418 encodes blood coagulation factor Xa  
 CC recognition sequence IEGR (AAR32109). GA is added immediately after  
 CC the CDS to form restriction enzyme NruI recognition site. A DNA  
 CC having inverted repeats of the first 11 nucleotides of the CDS is  
 CC ligated to form a palindrome sequence of 22 nucleotide pairs.

SQ Sequence 4 AA;

Query Match 100.0%; Score 20; DB 14; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
 ||||  
 Db 1 IEGR 4

RESULT 7

AAR39390

ID AAR39390 standard; Protein; 4 AA.

XX AAR39390;

XX 12-JAN-1994 (first entry)

XX Factor Xa cleavage site.

XX Rapid; protein isolation; calcium; presence; fusion protein;  
 XX affinity chromatography.

XX Synthetic.

XX Key Location/Qualifiers

FT Cleavage-site 1..4  
 /note= "factor Xa cleavage site"

XX WO9313211-A.

XX 08-JUL-1993.

XX 29-DEC-1992; 92MO-US11270.

XX 03-JAN-1992; 92US-0816679.

XX (OKLA-) OKLAHOMA MED RES FOUND.

XX Esmon CT, Morrissey JH, Rezaie A;

XX WPI; 1993-227327/28.

XX Fusion protein allowing rapid isolation of protein from soln. -  
 XX comprises desired protein and epitope recognised by monoclonal  
 XX antibody HPC-4

XX Example; Page 30; 42pp; English.

XX The sequence is that of the factor Xa cleavage site which was used  
 XX in the prepn. of a fusion protein comprising the epitope for the  
 XX Ca++ dependent monoclonal antibody HPC-4 and a protein to be  
 XX isolated. This fusion protein may be isolated using HPC-4-based  
 XX affinity chromatography, the protein can be recovered in a single  
 XX chromatographic step using immobilised HPC-4, and released from the  
 XX fusion protein by cleavage with factor Xa.

SQ Sequence 4 AA;

Query Match 100.0%; Score 20; DB 14; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
 ||||  
 Db 1 IEGR 4

RESULT 8

AAR54863

ID AAR54863 standard; peptide; 4 AA.

XX AAR54863;

XX 09-JAN-1995 (first entry)

XX Factor Xa cleavage site.

XX Human calcitonin; recombinant production; fusion protein;  
 XX Chloramphenicol acetyl transferase; Factor Xa cleavage site;  
 XX cell-free protein synthesis system.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Cleavage-site 1..4  
 /label= Factor\_Xa\_recognition\_site  
 /note= "fused to N-terminus of hCT"

XX JP06098790-A.

XX 12-APR-1994.

XX 04-MAR-1992; 92JP-0081480.

XX 04-MAR-1992; 92JP-0081480.

XX (KOBM ) KOBE STEEL LTD.

XX WPI; 1994-155930/19.

XX Prodn. of polypeptide in non-cellular protein synthesis system -  
 XX by constructing a fusion gene with chloramphenicol acetyl  
 XX transferase gene and opt. cleaving the polypeptide from the  
 XX resultant fusion protein.

XX Example 1; Page 4; 7pp; Japanese.

XX A synthetic coding sequence for human calcitonin was constructed  
 XX which coded for hCT (1-32) with an additional amidated C-terminal  
 XX Gly residue. The coding sequence was fused immediately downstream of  
 XX a region coding for a Factor Xa cleavage site (Ile-Glu-Gly-Arg),  
 XX which was itself in-frame, within the chloramphenicol acetyl  
 XX transferase (CAT) gene and downstream of a lac or tac promoter.  
 XX Human CT can be recovered from the resulting CAT-(IEGR)-hCT fusion  
 XX protein by cleavage with factor Xa. The synthetic hCT could be  
 XX replaced by sequences coding for other useful polypeptides.

SQ Sequence 4 AA;

Query Match 100.0%; Score 20; DB 15; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
 ||||  
 Db 1 IEGR 4

RESULT 9

AAR48071

ID AAR48071 standard; peptide; 4 AA.

XX AAR48071;

XX AAR48071;

DT 13-JUL-1994 (first entry)  
 XX Biological protecting gp. contg. Factor Xa cleavage site.  
 DE  
 XX  
 XX C-terminal alpha-amide polypeptide; amidation; protecting group;  
 KW N-terminal alpha-acetyl polypeptide; acetylation;  
 KW recombinant multicopy fusion protein; interconnecting peptide;  
 KW intraconnecting peptide; Blood coagulation Factor Xa;  
 KW recognition site.  
 XX  
 OS Synthetic.  
 XX  
 XX Key Location/Qualifiers  
 FH Cleavage-site 1..4  
 FT /label= Factor\_Xa\_recognition\_site  
 XX  
 XX WO9401451-A.  
 PN  
 XX  
 PD 20-JAN-1994.  
 XX  
 XX 13-JUL-1993; 93WO-US06591.  
 XX  
 XX 13-JUL-1992; 92US-0912798.  
 PR  
 XX (BION-) BIONEERASKA INC.  
 PA  
 XX Coolidge TR, Holmquist B, Stout J, Wagner FW;  
 PI WPI; 1994-034983/04.  
 DR N-PSDB; AAQ55445.  
 XX  
 XX Terminal modification of recombinant single copy polypeptide - by  
 PT protecting, modifying and de-protecting polypeptide, e.g. fusion  
 PT protein, contg. biologically added protecting gp.  
 XX  
 PS Disclosure and Claim 10; Page 22; 102pp; English.  
 XX  
 XX In order to produce recombinant polypeptides with C-terminal  
 CC alpha-amide groups and/or N-terminal acetyl groups, fusion proteins  
 CC are formed in which the relevant terminus (termini) of the  
 CC polypeptide is protected by a biological protecting group. The  
 CC protecting group is a peptide or amino acid having at least one  
 CC cleavage site for its removal. The sequence AAR48071 represents a  
 CC suitable biological protecting group, i.e. a Factor Xa cleavage  
 CC site.  
 XX  
 XX Sequence 4 AA;  
 SQ  
 Query Match 100.0%; Score 20; DB 15; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IEGR 4  
 Db | | | |  
 1 IEGR 4  
 RESULT 10  
 AAR60504  
 ID AAR60504 standard; Peptide; 4 AA.  
 XX  
 XX AAR60504;  
 AC  
 XX 22-MAR-1995 (first entry)  
 DT  
 XX Factor Xa cleavage site.  
 DE  
 XX Serine protease; Factor-Xa; recognition site;  
 KW fusion protein cleavage; protein folding.  
 KW  
 XX Synthetic.  
 OS  
 XX WO9418227-A.  
 PN

XX 18-AUG-1994.  
 PD  
 XX 04-FEB-1994; 94WO-DK000054.  
 PF  
 XX 04-FEB-1993; 93DK-0000130.  
 PR 05-FEB-1993; 93DK-0000139.  
 PR 03-DEC-1993; 93WO-GB02492.  
 XX  
 XX (DENZ-) DENZYME APS.  
 FA  
 XX Etzerodt M, Holtet TL, Thogersen HC;  
 PI WPI; 1994-279681/34.  
 DR  
 XX Refolding of polypeptide molecules - using a cyclic process  
 PT involving denaturing and renaturing conditions to produce a  
 PT correctly folded prod  
 PT  
 XX Disclosure; Page 126; 202pp; English.  
 PS  
 XX Novel recognition sites for cattle Factor-Xa are given in  
 CC AAR60504 and AAR60506-12. Such sequences may replace the Factor-Xa  
 CC recognition sequence (AAR60503) in recombinant fusion proteins to  
 CC facilitate cleavage by a new serine protease (given in AAR60502).  
 XX  
 XX Sequence 4 AA;  
 SQ  
 Query Match 100.0%; Score 20; DB 15; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IEGR 4  
 Db | | | |  
 1 IEGR 4  
 RESULT 11  
 AAR60693  
 ID AAR60693 standard; Protein; 4 AA.  
 XX  
 XX AAR60693;  
 AC  
 XX 25-JUN-1995 (first entry)  
 DT  
 XX H. halobium bacteriorhodopsin protease cleavage site.  
 DE  
 XX Halobacteria; expression.  
 KW  
 XX Halobacterium halobium strain R1.  
 OS  
 XX WO9421789-A.  
 PN  
 XX 29-SEP-1994.  
 PD  
 XX 28-FEB-1994; 94WO-US02388.  
 PF  
 XX 25-MAR-1993; 93US-0038662.  
 PR  
 XX (REGC ) UNIV CALIFORNIA.  
 PA  
 XX Betlach MC, Turner GJ;  
 PI WPI; 1994-317010/39.  
 DR  
 XX Expression of heterologous proteins in halo-bacteria - using  
 PT regulatory and stop sequences from halo-bacteria, pref. the  
 PT bacterio-rhodopsin gene.  
 PT  
 XX Disclosure; Page 31; 118pp; English.  
 PS  
 XX The sequence is that of the Halobacterium halobium bacteriorhodopsin  
 CC protease cleavage site. It is used to exemplify a new expression  
 CC

CC vector for producing heterologous polypeptides in a halobacterial host.  
 XX See also AAR60691-9.  
 SQ Sequence 4 AA;

Query Match 100.0%; Score 20; DB 15; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
 ||||  
 Db 1 IEGR 4

RESULT 12  
 AAR86901  
 ID AAR86901 standard; Peptide; 4 AA.  
 XX  
 AC AAR86901;  
 XX

DT 16-MAY-1996 (first entry)  
 XX

DE Blood coagulation factor Xa selective cleavage site.  
 XX

KW GRB; growth factor receptor bound; tyrosine kinase; regulation;  
 cell growth; cellular metabolism; screening; signal transduction;  
 KW cancer; diabetes; CORT technique; cloning of receptor targets.  
 XX

OS Homo sapiens.  
 XX

XX WO9524426-A1.  
 XX

PD 14-SEP-1995.  
 XX

PF 13-MAR-1995; 95WO-US03385.  
 XX

PR 11-MAR-1994; 94US-0208887.  
 XX

PA (UWNY ) UNIV NEW YORK STATE.  
 XX

PI Margolis BL, Schlessinger J, Skolnik EY;  
 XX

DR WPI; 1995-328235/42.  
 XX

PT DNA encoding tyrosine kinase-binding proteins - used to screen  
 agents capable of modulating cell growth or cellular metabolism

PS Disclosure; Page 46; 215pp; English.  
 XX

CC AAR86901 and AAR86902 are selective cleavage sites which can be  
 incorporated into a plasmid used for the recombinant production of GRB  
 proteins. Using a new cloning technique, CORT (cloning of receptor  
 targets) several new TK binding proteins were isolated. Growth factor  
 receptor bound proteins GRB-1, GRB-2, GRB-3, GRB-4, GRB-7 and GRB-10 were  
 isolated using this method. The proteins bind to a tyrosine-  
 phosphorylated domain of a eukaryotic TK. GRB proteins can be used for  
 screening agents which are capable of modulating cell growth that  
 occurs via signal transduction through TKs. Such agents can be used to  
 prevent or inhibit cell growth or to counteract tumour development.  
 CC GRB proteins are also useful for identifying susceptibility to  
 diseases associated with alterations in cellular metabolism mediated by  
 CC TK pathways e.g. cancer and diabetes.  
 XX

SQ Sequence 4 AA;

Query Match 100.0%; Score 20; DB 16; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
 ||||  
 Db 1 IEGR 4

RESULT 13  
 AAR88219

ID AAR88219 standard; peptide; 4 AA.  
 XX

AC AAR88219;  
 XX

DT 13-JUN-1996 (first entry)  
 XX

DE Lactoferrisin antibacterial peptide associated protease fragment.  
 XX

KW Recombinant vector; generic; lactoferrisin; antibacterial peptide;  
 regulatory sequence; tac; promoter; shuttle; vector; pGEX2; GAL1;  
 KW pKOM2; Rous Sarcoma Virus; long terminal repeat; pRSVNot;  
 KW lactoferricin; protease fragment.  
 XX

OS Synthetic.  
 XX

PN JP07274970-A.  
 XX

PD 24-OCT-1995.  
 XX

PF 01-APR-1994; 94JP-0085244.  
 XX

PR 01-APR-1994; 94JP-0085244.  
 XX

PA (MORG ) MORINAGA MILK IND CO LTD.  
 XX

DR WPI; 1995-399338/51.  
 XX

PT Recombinant vector contg. lactoferrisin gene - used to prepare an  
 antibacterial peptide

PS Example 1; Page 12; 18pp; Japanese.  
 XX

CC A recombinant vector in which a DNA sequence encoding at least the  
 generic lactoferrisin antibacterial peptide AAR88216 is inserted,  
 CC pref. downstream of the vector's regulatory sequence, is claimed.  
 CC The protease fragment AAR88219 and the DNA sequences AAT08774-79 were  
 used in the construction of such a vector, where the regulatory  
 CC sequence is the tac promoter from shuttle vector pGEX2, the GAL1  
 CC promoter from vector pKOM2 or Rous Sarcoma Virus long terminal  
 CC repeat from vector pRSVNot.  
 XX

SQ Sequence 4 AA;

Query Match 100.0%; Score 20; DB 16; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
 ||||  
 Db 1 IEGR 4

RESULT 14

AAR83116  
 ID AAR83116 standard; Peptide; 4 AA.  
 XX

AC AAR83116;  
 XX

DT 29-APR-1996 (first entry)  
 XX

DE Factor-Xa recognition sequence.  
 XX

KW Calcitonin; transgenic animal; milk; Paget disease; hypercalcaemia;  
 osteoporosis; fusion protein cleavage; Factor-Xa.  
 XX

OS Synthetic.  
 XX

PN WO9527782-A1.  
 XX

PD 19-OCT-1995.  
 XX

XX 05-APR-1995; 95WO-GB00769.  
 XX 08-APR-1994; 94GB-0006974.  
 XX (PPLT-) PPL THERAPEUTICS SCOTLAND LTD.  
 XX Cottingham IR, Garner I;  
 XX WPI; 1995-366387/47.  
 XX Peptide, pref. calcitonin, prodn. in transgenic non-human mammal -  
 PT by expressing fusion protein in the milk followed by cleavage and  
 PT sepn., used to treat Paget's disease, hypercalcaemic shock etc.  
 XX Claim 13; Page 21; 26pp; English.  
 XX The recognition sequence (AAR83116) for Factor-Xa may be included in  
 CC the linker segment of a fusion protein, comprising e.g. calcitonin  
 CC and alpha-lactalbumin, produced in the milk of a transgenic mammal.  
 CC Cleavage of the fusion protein with Factor-Xa yields alpha-  
 CC lactalbumin, which is removed by an affinity method, and purified  
 CC calcitonin of therapeutic appin. Alternatively, the linker  
 CC comprises the activation peptide (AAR83118) of trypsinogen, which  
 CC includes the recognition sequence (AAR83117) for enterokinase.  
 XX SQ Sequence 4 AA;  
 Query Match 100.0%; Score 20; DB 16; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IEGR 4  
 Db 1 IEGR 4  
 RESULT 15  
 AAR73943  
 ID AAR73943 standard; Peptide; 4 AA.  
 XX AC AAR73943;  
 XX 24-OCT-1995 (first entry)  
 XX Endotoxin assay peptide.  
 XX Endotoxin; assay; Limulus amoebocyte lysate; cascade reaction.  
 XX Synthetic.  
 XX Key Location/Qualifiers  
 FH Modified-site 1 /note= "N-terminal Boc group"  
 FT Modified-site 4 /note= "C-terminal pNa group"  
 FT EP649021-A.  
 XX PN 19-APR-1995.  
 XX 30-SEP-1994; 94EP-0115442.  
 XX 30-SEP-1993; 93JP-0265479.  
 XX (SEK ) SEIKAGAKU KOGYO CO LTD.  
 XX Oda T, Tamura H, Tanaka S;  
 XX WPI; 1995-148858/20.  
 XX Limulus amoebocyte lysate reagent for endotoxin determ. - contg.  
 PT alkyl glucoside to inhibit activation of factor G by

PT (1-3)-beta-D-glucan  
 XX Disclosure; Page 15; 20pp; English.  
 XX The peptides given in AAR73942-44 are used in endotoxin-specific  
 CC assays in which clotting enzymes, formed in cascade reactions  
 CC initiated by addition of endotoxin, hydrolyze an amide bond in  
 CC the peptides, thereby liberating chromogenic p-nitroaniline.  
 XX SQ Sequence 4 AA;  
 Query Match 100.0%; Score 20; DB 16; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IEGR 4  
 Db 1 IEGR 4  
 Search completed: April 13, 2003, 03:20:34  
 Job time : 4.22278 secs





GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2003, 03:20:41 ; Search time 0.0997506 Seconds  
(without alignments)  
1179.859 Million cell updates/sec

Title: US-10-023-888-22

Perfect score: 20

Sequence: 1 IEGR 4

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued Patents AA.\*  
2: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/PTUS\_COMB.pep.\*  
7: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	4	1	US-09-377-465A-4
2	20	100.0	4	1	US-07-816-679A-6
3	20	100.0	4	1	US-08-030-731A-2
4	20	100.0	4	1	US-07-854-596B-71
5	20	100.0	4	1	US-08-160-670A-2
6	20	100.0	4	1	US-08-176-500-19
7	20	100.0	4	1	US-08-243-082-20
8	20	100.0	4	1	US-08-170-095B-16
9	20	100.0	4	1	US-08-240-712-2
10	20	100.0	4	1	US-08-240-712-11
11	20	100.0	4	1	US-08-167-035-15
12	20	100.0	4	1	US-08-471-052A-19
13	20	100.0	4	1	US-08-255-272-20
14	20	100.0	4	1	US-08-294-434-3
15	20	100.0	4	1	US-08-468-674B-89
16	20	100.0	4	1	US-08-313-553-4
17	20	100.0	4	1	US-08-110-300A-20
18	20	100.0	4	1	US-08-457-166-3
19	20	100.0	4	1	US-08-396-866-16
20	20	100.0	4	1	US-08-208-887A-15
21	20	100.0	4	1	US-08-313-899A-2
22	20	100.0	4	1	US-08-443-890-2
23	20	100.0	4	1	US-08-443-890-11
24	20	100.0	4	1	US-08-469-486-38
25	20	100.0	4	1	US-08-189-331-19
26	20	100.0	4	1	US-08-153-799-23
27	20	100.0	4	1	US-08-780-571-89

28	20	100.0	4	1	US-08-621-081A-12
29	20	100.0	4	2	US-08-115-519-2
30	20	100.0	4	2	US-08-754-431A-9
31	20	100.0	4	2	US-08-235-515A-19
32	20	100.0	4	2	US-08-471-939-19
33	20	100.0	4	2	US-08-471-800-19
34	20	100.0	4	2	US-08-539-005-15
35	20	100.0	4	2	US-08-203-662-1
36	20	100.0	4	2	US-08-476-014-2
37	20	100.0	4	2	US-07-733-095B-2
38	20	100.0	4	2	US-08-469-658-38
39	20	100.0	4	2	US-07-982-064-6
40	20	100.0	4	2	US-08-471-068-19
41	20	100.0	4	2	US-08-886-642-20
42	20	100.0	4	2	US-08-595-868C-18
43	20	100.0	4	3	US-08-752-892-1
44	20	100.0	4	3	US-08-767-993-4
45	20	100.0	4	3	US-08-722-126A-11

## ALIGNMENTS

RESULT 1  
US-09-377-465A-4  
; Sequence 4, Application US/09377465A  
; Patent No. H002021  
; GENERAL INFORMATION:  
; APPLICANT: Hoskins, JoAnn  
; APPLICANT: Jaskunas, Stanley R  
; APPLICANT: Zhao, Genshi  
; APPLICANT: Rockey, Pamela K  
; TITLE OF INVENTION: NOVEL PENICILLIN BINDING PROTEIN FROM STREPTOCOCCUS  
; FILE REFERENCE: PNEUMONIAE  
; Patent No. H002021  
; CURRENT APPLICATION NUMBER: US/09377,465A  
; CURRENT FILING DATE: 1999-08-19  
; PRIOR APPLICATION NUMBER: 60/100,887  
; PRIOR FILING DATE: 1998-09-23  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Unknown Organism  
; FEATURE:  
; OTHER INFORMATION: Description of Unknown Organism: Thrombin Cleavage  
; OTHER INFORMATION: Site  
US-09-377-465A-4

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 1.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
Db 1 IEGR 4

RESULT 2  
US-07-816-679A-6  
; Sequence 6, Application US/07816679A  
; Patent No. 5298599  
; GENERAL INFORMATION:  
; APPLICANT: Rezale, Alireza  
; APPLICANT: Esmon, Charles T.  
; APPLICANT: Morrissey, James H.  
; TITLE OF INVENTION: Expression and Purification of  
; TITLE OF INVENTION: Recombinant Soluble Tissue Factor  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kilpatrick & Cody

```
STREET: 100 Peachtree Street, Suite 3200
CITY: Atlanta
STATE: Georgia
COUNTRY: U.S.
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/07/816,679A
  FILING DATE: 19920103
  CLASSIFICATION: 435
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 07/730040
    FILING DATE: 12-JUL-1991
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 07/292447
    FILING DATE: 30-DEC-1988
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 07/683682
    FILING DATE: 10-APR-1991
  ATTORNEY/AGENT INFORMATION:
    NAME: Pabst, Patrea L.
    REGISTRATION NUMBER: 31,284
    REFERENCE/DOCKET NUMBER: OMR130
    TELEPHONE: 404-572-6508
    TELEFAX: 404-572-6555
  INFORMATION FOR SEQ ID NO: 6:
    LENGTH: 4 amino acids
    TYPE: AMINO ACID
    STRANDEDNESS: single
    TOPOLOGY: linear
    MOLECULE TYPE: peptide
    HYPOTHETICAL: NO
    ANTI-SENSE: NO
    FEATURE:
      NAME/KEY: Cleavage-site
      LOCATION: 1..4
      OTHER INFORMATION: /note="Factor Xa Cleavage Site"
US-07-816-679A-6
Query Match 100.0%; Score 20; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.9e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 IEGR 4
Db 1 IEGR 4
RESULT 3
US-08-030-731A-2
Sequence 2, Application US/08030731A
Patent No. 5426036
GENERAL INFORMATION:
  APPLICANT: Koller, Klaus-Peter
  APPLICANT: Riess, Guenther Johannes
  APPLICANT: Uhlmann, Eugen
  APPLICANT: Wallmeier, Holger
  TITLE OF INVENTION: Processes for the Preparation of Foreign
  TITLE OF INVENTION: Proteins in Streptomycetes
  NUMBER OF SEQUENCES: 48
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Dunner
    STREET: 1300 I Street, N.W., Suite 700
    CITY: Washington
    STATE: D.C.
    COUNTRY: USA
STREET: 100 Peachtree Street, Suite 3200
CITY: Atlanta
STATE: Georgia
COUNTRY: U.S.
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/07/816,679A
  FILING DATE: 19920103
  CLASSIFICATION: 435
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 07/730040
    FILING DATE: 12-JUL-1991
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 07/292447
    FILING DATE: 30-DEC-1988
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 07/683682
    FILING DATE: 10-APR-1991
  ATTORNEY/AGENT INFORMATION:
    NAME: Pabst, Patrea L.
    REGISTRATION NUMBER: 31,284
    REFERENCE/DOCKET NUMBER: OMR130
    TELEPHONE: 404-572-6508
    TELEFAX: 404-572-6555
  INFORMATION FOR SEQ ID NO: 6:
    LENGTH: 4 amino acids
    TYPE: AMINO ACID
    STRANDEDNESS: single
    TOPOLOGY: linear
    MOLECULE TYPE: peptide
    HYPOTHETICAL: NO
    ANTI-SENSE: NO
    FEATURE:
      NAME/KEY: Cleavage-site
      LOCATION: 1..4
      OTHER INFORMATION: /note="Factor Xa Cleavage Site"
US-07-816-679A-6
Query Match 100.0%; Score 20; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.9e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 IEGR 4
Db 1 IEGR 4
RESULT 3
US-08-030-731A-2
Sequence 2, Application US/08030731A
Patent No. 5426036
GENERAL INFORMATION:
  APPLICANT: Koller, Klaus-Peter
  APPLICANT: Riess, Guenther Johannes
  APPLICANT: Uhlmann, Eugen
  APPLICANT: Wallmeier, Holger
  TITLE OF INVENTION: Processes for the Preparation of Foreign
  TITLE OF INVENTION: Proteins in Streptomycetes
  NUMBER OF SEQUENCES: 48
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Dunner
    STREET: 1300 I Street, N.W., Suite 700
    CITY: Washington
    STATE: D.C.
    COUNTRY: USA
STREET: 100 Peachtree Street, Suite 3200
CITY: Atlanta
STATE: Georgia
COUNTRY: U.S.
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/030,731A
  FILING DATE: 12-MAR-1993
  CLASSIFICATION: 435
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 07/189,840
    FILING DATE: 03-MAY-1988
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 07/430,622
    FILING DATE: 01-NOV-1989
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 07/687,610
    FILING DATE: 19-APR-1991
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 07/735,757
    FILING DATE: 29-JUL-1991
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: DE P 37 14 866.4
    FILING DATE: 05-MAY-1987
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: DE P 38 37 273.8
    FILING DATE: 03-NOV-1988
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: DE P 39 27 449.7
    FILING DATE: 19-AUG-1989
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: DE P 40 12 818.0
    FILING DATE: 21-APR-1990
  ATTORNEY/AGENT INFORMATION:
    NAME: Kirschner Michael K.
    REGISTRATION NUMBER: 34,851
    REFERENCE/DOCKET NUMBER: 02481-0593-02000
    TELEPHONE: 202-408-4000
    TELEFAX: 202-408-4400
  INFORMATION FOR SEQ ID NO: 2:
    LENGTH: 4 amino acids
    TYPE: amino acid
    TOPOLOGY: unknown
    MOLECULE TYPE: peptide
US-08-030-731A-2
Query Match 100.0%; Score 20; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.9e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 IEGR 4
Db 1 IEGR 4
RESULT 4
US-854-596B-71
Sequence 71, Application US/07854596B
Patent No. 5434073
GENERAL INFORMATION:
  APPLICANT: Dawson, Keith M
  APPLICANT: Hunter, Michael G
  APPLICANT: Czaplewski, Lloyd G
  TITLE OF INVENTION: Proteins and nucleic acids
  NUMBER OF SEQUENCES: 73
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Dr. John J. McDonnell
    STREET: Ten South Wacker Drive, Suite 3000
    CITY: Chicago
    STATE: IL
```

; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA: US/07/854,596B  
; APPLICATION NUMBER: US/07/854,596B  
; FILING DATE: 03-JUN-1992  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McDonnell, John J.  
; REGISTRATION NUMBER: 26,949  
; REFERENCE/DOCKET NUMBER: 92,337  
; TELEPHONE: 312-715-1000  
; TELEFAX: 312-715-1234  
; TELEX: 910-221-5317  
; INFORMATION FOR SEQ ID NO: 71:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 1..4  
; OTHER INFORMATION: /note= "Factor Xa cleavable"  
; OTHER INFORMATION: linker"  
; US-07-854-596B-71

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 1.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
Db 1 IEGR 4

RESULT 5  
; Sequence 2, Application US/08160670A  
; Patent No. 5449758  
; GENERAL INFORMATION:  
; APPLICANT: Hartley, James L.  
; TITLE OF INVENTION: Protein Size Marker Ladder  
; NUMBER OF SEQUENCES: 49  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
; STREET: 1100 New York Avenue, Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/160,670A  
; FILING DATE: 12/2/93  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Esmond, Robert W.  
; REGISTRATION NUMBER: 32,893  
; REFERENCE/DOCKET NUMBER: 0942.2580000  
; TELEPHONE: (202) 371-2600  
; TELEFAX: (202) 371-2540  
; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4 amino acids  
; TYPE: amino acid  
; TOPOLOGY: both  
; US-08-160-670A-2

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 1.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
Db 1 IEGR 4

RESULT 6  
; US-08-176-500-19  
; Sequence 19, Application US/08176500  
; Patent No. 5498538  
; GENERAL INFORMATION:  
; APPLICANT: Kay, B. K.  
; TITLE OF INVENTION: Totally Synthetic Affinity Reagents  
; NUMBER OF SEQUENCES: 141  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/176,500  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/013,416  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mistock, S. Leslie  
; REGISTRATION NUMBER: 18,872  
; REFERENCE/DOCKET NUMBER: 1101-143  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212 790-9090  
; TELEFAX: 212 869-8864/9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 19:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
; US-08-176-500-19

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 1.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
Db 1 IEGR 4

RESULT 7  
; US-08-243-082-20  
; Sequence 20, Application US/08243082  
; Patent No. 5506120

;; GENERAL INFORMATION:  
;; APPLICANT: YAMAMOTO, Hiroaki  
;; APPLICANT: YAMASHITA, Kunihiko  
;; TITLE OF INVENTION: METHOD OF PRODUCING PEPTIDES OR  
;; TITLE OF INVENTION: PROTEINS  
;; NUMBER OF SEQUENCES: 26  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Spencer, Frank & Schneider  
;; STREET: 1111 Nineteenth Street, N.W.  
;; CITY: Washington  
;; STATE: D.C.  
;; COUNTRY: U.S.A.  
;; ZIP: 20036  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/243,082  
;; FILING DATE:  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US/07/853,754  
;; FILING DATE: 05-JUN-1992  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Schneller, John W.  
;; REGISTRATION NUMBER: 26,031  
;; REFERENCE/DOCKET NUMBER: KUWAT 0010  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (202) 828-8000  
;; TELEFAX: (202) 828-8038  
;; TELEX: SPENCER 64267  
;; INFORMATION FOR SEQ ID NO: 20:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 4 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; US-08-243-082-20  
  
Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 1.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 IEGR 4  
Db 1 IEGR 4  
  
RESULT 8  
US-08-170-095B-16  
; Sequence 16, Application US/08170095B  
; Patent No. 5563254  
; GENERAL INFORMATION:  
; APPLICANT: Hoffman, Stephen J.  
; APPLICANT: Nagai, Kiyoshi  
; TITLE OF INVENTION: Blood Substitutes  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Somatogen, Inc.  
; STREET: 2545 Central Avenue  
; CITY: Boulder  
; STATE: Colorado  
; ZIP: 80301  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage  
; COMPUTER: Apple Macintosh  
; OPERATING SYSTEM: System 7.0.1  
; SOFTWARE: Microsoft Word 5.0a  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/170,095B  
; FILING DATE: December 20, 1993

;; CLASSIFICATION: 530  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: No. 5563254ak, Henry P.  
;; REGISTRATION NUMBER: 33200  
;; REFERENCE/DOCKET NUMBER: Hoffman 2A/CONT2  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 303-541-3322  
;; TELEFAX: 303-444-3013  
;; INFORMATION FOR SEQ ID NO: 16:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 4  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: unknown to applicant  
;; MOLECULE TYPE: peptide  
;; HYPOTHETICAL: no  
;; US-08-170-095B-16  
  
Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 1.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 IEGR 4  
Db 1 IEGR 4  
  
RESULT 9  
US-08-240-712-2  
; Sequence 2, Application US/08240712  
; Patent No. 5599907  
; GENERAL INFORMATION:  
; APPLICANT: ANDERSON, DAVID C.  
; APPLICANT: MATHEWS, ANTONY JAMES  
; APPLICANT: STETLER, GARY L.  
; TITLE OF INVENTION: PRODUCTION AND USE OF MULTIMERIC  
; TITLE OF INVENTION: HEMOGLOBINS  
; NUMBER OF SEQUENCES: 35  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brody and Neimark  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/240,712  
; FILING DATE: 09-MAY-1994  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/09752  
; FILING DATE: 13-MAY-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: COOPER, IVER P  
; REGISTRATION NUMBER: 28,005  
; REFERENCE/DOCKET NUMBER: ANDERSON-6  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; TELEX: 248633  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
;; US-08-240-712-2

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 1.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
||||  
Db 1 IEGR 4

RESULT 10  
US-08-240-712-11  
; Sequence 11, Application US/08240712  
; Patent No. 559907  
; GENERAL INFORMATION:  
; APPLICANT: ANDERSON, DAVID C.  
; APPLICANT: MATHEWS, ANTONY JAMES  
; APPLICANT: STETTLER, GARY L.  
; TITLE OF INVENTION: PRODUCTION AND USE OF MULTIMERIC  
; TITLE OF INVENTION: HEMOGLOBINS  
; NUMBER OF SEQUENCES: 35  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Broadway and Neimark  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/240,712  
; FILING DATE: 09-MAY-1994  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/09752  
; FILING DATE: 13-MAY-1993  
; NAME: COOPER, IVER P  
; ATTORNEY/AGENT INFORMATION:  
; REGISTRATION NUMBER: 28,005  
; REFERENCE/DOCKET NUMBER: ANDERSON-6  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; TELEX: 248633

INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-240-712-11

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 1.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
||||  
Db 1 IEGR 4

RESULT 11  
US-08-167-035-15  
; Sequence 15, Application US/08167035  
; Patent No. 5618691  
; GENERAL INFORMATION:  
; APPLICANT: Schlensing, Joseph  
; APPLICANT: Skolnick, Edward Y.

APPLICANT: Margolis, Benjamin L.  
; TITLE OF INVENTION: NOVEL EXPRESSION CLONING METHOD FOR  
; TITLE OF INVENTION: IDENTIFYING TARGET PROTEINS FOR EUKARYOTIC TYROSINE  
; TITLE OF INVENTION: KINASES AND NOVEL TARGET PROTEINS  
; NUMBER OF SEQUENCES: 50  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: 10036-2711  
; ZIP: 10036-2711  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/167,035  
; FILING DATE: 16-DEC-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7683-062  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-9741/8864  
; TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
US-08-167-035-15

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 1.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
||||  
Db 1 IEGR 4

RESULT 12  
US-08-471-052A-19  
; Sequence 19, Application US/08471052A  
; Patent No. 5625033  
; GENERAL INFORMATION:  
; APPLICANT: Ray, B. K.  
; APPLICANT: Fowlkes, D. M.  
; TITLE OF INVENTION: Totally Synthetic Affinity Reagents  
; NUMBER OF SEQUENCES: 166  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/471,052A  
; FILING DATE: 06-JUNE-1995  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mistrock, S. Leslie

```

; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-179
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-471-052A-19

Query Match      100.0%; Score 20; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.9e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4
   ||||
DB 1 IEGR 4

RESULT 13
US-08-255-272-20
; Sequence 20, Application US/08255272
; Patent No. 5627268
; GENERAL INFORMATION:
; APPLICANT: Kumar, Ramesh
; APPLICANT: Sharma, Ajay
; APPLICANT: Khoury-Christianson, Anastasia
; APPLICANT: M.
; TITLE OF INVENTION: Production of Therapeutic Peptides in
; TITLE OF INVENTION: Transgenic Animals as a Fusion with Hemoglobin
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30742
; REFERENCE/DOCKET NUMBER: 6794-032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-255-272-20

Query Match      100.0%; Score 20; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.9e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4
   ||||
DB 1 IEGR 4

RESULT 14
US-08-294-434-3
; Sequence 3, Application US/08294434
; Patent No. 5635371
; GENERAL INFORMATION:
; APPLICANT: Stout, Jay
; APPLICANT: Wagner, Fred W.
; APPLICANT: Coolidge, Thomas R.
; APPLICANT: Holmquist, Barton
; TITLE OF INVENTION: CHEMICAL METHOD FOR SELECTIVE
; TITLE OF INVENTION: MODIFICATION OF THE N- AND/OR C-TERMINAL AMINO ACID
; TITLE OF INVENTION: ALPHA-CARBON REACTIVE GROUP OF A RECOMBINANT
; TITLE OF INVENTION: POLYPEPTIDE OR A PORTION THEREOF
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant & Gould
; STREET: 3100 No. 5635371 West Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/294,434
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/912,798
; FILING DATE: 13-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Nelson, Albin J.
; REGISTRATION NUMBER: 28,650
; REFERENCE/DOCKET NUMBER: 8648.29-US01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-332-5300
; TELEFAX: 612-332-9081
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-294-434-3

Query Match      100.0%; Score 20; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.9e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4
   ||||
DB 1 IEGR 4

RESULT 15
US-08-468-674B-89
; Sequence 89, Application US/08468674B
; Patent No. 5639642
; GENERAL INFORMATION:
; APPLICANT: Kjeldsen, Thomas B
; APPLICANT: Vad, Knud
; TITLE OF INVENTION: SYNTHETIC LEADERS PEPTIDE SEQUENCES
; NUMBER OF SEQUENCES: 89
; CORRESPONDENCE ADDRESS:
```

ADDRESSEE: No. 5639642o No. 5639642disk of No. 5639642th America, Inc.  
STREET: 405 Lexington Avenue, 64th Floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Tape  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/468,674B  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/282,852  
FILING DATE: 29-JUL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lambiris, Elias J.  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 4085.220-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 89:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Peptide  
US-08-468-674B-89

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 1.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
|  
|  
|  
|  
Db 1 IEGR 4

Search completed: April 13, 2003, 03:29:22  
Job time : 1.09975 secs





GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2003, 03:22:54 ; Search time 0.113051 Seconds  
(without alignments)  
2163.137 Million cell updates/sec

Title: US-10-023-888-22  
Perfect score: 20  
Sequence: 1 IEGR 4

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 248812 seqs, 6113604 residues

Total number of hits satisfying chosen parameters: 248812

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	4	9	US-09-858-332-8
2	20	100.0	4	9	US-09-832-355A-64
3	20	100.0	4	10	US-09-040-518-6
4	20	100.0	4	10	US-09-998-831-21
5	20	100.0	5	9	US-09-770-102A-51
6	20	100.0	5	9	US-09-259-658-23
7	20	100.0	5	10	US-09-970-308-5
8	20	100.0	5	10	US-09-904-117-9
9	20	100.0	6	9	US-09-899-235-29
10	20	100.0	7	9	US-10-150-262-11
11	20	100.0	8	12	US-10-066-209-6
12	20	100.0	8	12	US-10-066-209-8
13	20	100.0	10	10	US-09-809-517A-2
14	20	100.0	11	9	US-10-092-908-30
15	20	100.0	11	9	US-10-092-908-46
16	20	100.0	15	9	US-09-880-132-62
17	20	100.0	15	10	US-09-880-149-62
18	20	100.0	16	9	US-10-067-790-40
19	20	100.0	16	9	US-10-067-892-40
					Sequence 8, Appli
					Sequence 64, Appli
					Sequence 6, Appli
					Sequence 21, Appli
					Sequence 51, Appli
					Sequence 23, Appli
					Sequence 5, Appli
					Sequence 29, Appli
					Sequence 11, Appli
					Sequence 8, Appli
					Sequence 2, Appli
					Sequence 30, Appli
					Sequence 46, Appli
					Sequence 62, Appli
					Sequence 62, Appli
					Sequence 40, Appli

20	100.0	16	9	US-09-539-382-40	Sequence 40, Appli	
21	20	100.0	16	9	US-10-067-893-40	Sequence 40, Appli
22	20	100.0	20	10	US-09-205-658-381	Sequence 281, Appli
23	20	100.0	21	9	US-10-092-908-47	Sequence 47, Appli
24	20	100.0	22	9	US-09-331-631A-16	Sequence 16, Appli
25	20	100.0	29	10	US-09-864-761-48408	Sequence 48408, A
26	20	100.0	31	9	US-10-092-908-48	Sequence 48, Appli
27	20	100.0	37	10	US-09-205-658-257	Sequence 257, Appli
28	20	100.0	41	9	US-10-092-908-49	Sequence 49, Appli
29	20	100.0	50	10	US-09-957-607-2	Sequence 2, Appli
30	20	100.0	52	9	US-10-091-572-242	Sequence 242, Appli
31	20	100.0	64	9	US-09-796-692-1068	Sequence 1068, Ap
32	20	100.0	64	9	US-09-796-692-1072	Sequence 1072, Ap
33	20	100.0	64	9	US-09-796-692-1080	Sequence 1080, Ap
34	20	100.0	64	9	US-09-796-692-1091	Sequence 1091, Ap
35	20	100.0	64	9	US-09-796-692-1565	Sequence 1565, Ap
36	20	100.0	64	9	US-09-796-692-1846	Sequence 1846, Ap
37	20	100.0	64	9	US-09-796-692-1738	Sequence 1738, Ap
38	20	100.0	76	9	US-10-104-755-79	Sequence 79, Appli
39	20	100.0	76	9	US-10-104-755-81	Sequence 81, Appli
40	20	100.0	76	9	US-10-104-755-83	Sequence 83, Appli
41	20	100.0	76	9	US-10-125-540-373	Sequence 373, Appli
42	20	100.0	76	10	US-09-764-870-373	Sequence 373, Appli
43	20	100.0	78	10	US-09-864-761-38379	Sequence 38379, A
44	20	100.0	86	9	US-09-884-767A-9	Sequence 9, Appli
45	20	100.0	92	9	US-09-738-626-3898	Sequence 3898, Ap

## ALIGNMENTS

RESULT 1

US-09-858-332-8  
; Sequence 8, Application US/09858332  
; Patent No. US20020164718A1  
; GENERAL INFORMATION:  
; APPLICANT: Tchega, Grigory S.  
; APPLICANT: Jomhadze, George  
; TITLE OF INVENTION: Metal Ion Affinity Tags and Methods for  
; FILE REFERENCE: Using the Same  
; CURRENT APPLICATION NUMBER: US/09/858,332  
; CURRENT FILING DATE: 2002-07-02  
; PRIOR APPLICATION NUMBER: 09/404,017  
; PRIOR FILING DATE: 1999-09-23  
; PRIOR APPLICATION NUMBER: 60/101,867  
; PRIOR FILING DATE: 1998-09-25  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: synthetic peptide  
US-09-858-332-8

Query Match 100.0%; Score 20; DB 9; Length 4;  
Best Local Similarity 100.0%; Pred. No. 2.2e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4

Db 1 IEGR 4

## RESULT 2

US-09-832-355A-64  
; Sequence 64, Application US/09832355A  
; Publication No. US20030027751A1  
; GENERAL INFORMATION:  
; APPLICANT: Kovesdi, Imre  
; APPLICANT: Kessler, Paul

; TITLE OF INVENTION: VEGF FUSION PROTEINS  
 ; FILE REFERENCE: 205654  
 ; CURRENT APPLICATION NUMBER: US/09/832,355A  
 ; CURRENT FILING DATE: 2001-04-10  
 ; NUMBER OF SEQ ID NOS: 126  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 64  
 ; LENGTH: 4  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC  
 ; OTHER INFORMATION: PEPTIDE  
 ; US-09-998-831-21

Query Match 100.0%; Score 20; DB 10; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
 ||||  
 Db 1 IEGR 4

RESULT 5  
 US-09-770-102A-51  
 ; Sequence 51, Application US/09770102A  
 ; Publication No. US20020197606A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cyclacel  
 ; TITLE OF INVENTION: Compositions and Methods for Monitoring the Modification of MO  
 ; FILE REFERENCE: 10069/1062  
 ; CURRENT APPLICATION NUMBER: US/09/770,102A  
 ; CURRENT FILING DATE: 2001-01-25  
 ; PRIOR APPLICATION NUMBER: US 60/179283  
 ; PRIOR FILING DATE: 2000-01-31  
 ; NUMBER OF SEQ ID NOS: 57  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 51  
 ; LENGTH: 5  
 ; TYPE: PRT  
 ; ORGANISM: Unknown  
 ; FEATURE:  
 ; OTHER INFORMATION: Cleavage site  
 ; NAME/KEY: SITE  
 ; LOCATION: (1)..(5)  
 ; OTHER INFORMATION: Cleavage site, Xaa is any amino acid  
 ; NAME/KEY: MISC FEATURE  
 ; LOCATION: (1)..(5)  
 ; OTHER INFORMATION: X at position 5 can be any amino acid  
 ; US-09-770-102A-51

Query Match 100.0%; Score 20; DB 10; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
 ||||  
 Db 1 IEGR 4

RESULT 6  
 US-09-259-658-23  
 ; Sequence 23, Application US/09259658  
 ; Publication No. US20030032034A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Colyer  
 ; APPLICANT: Craig  
 ; APPLICANT: Maschio  
 ; APPLICANT: Mezna  
 ; TITLE OF INVENTION: Compositions And Methods For Monitoring The  
 ; FILE REFERENCE: Modification State Of A Pair Of Polypeptides  
 ; FILE REFERENCE: colyer 4256/79245  
 ; CURRENT APPLICATION NUMBER: US/09/259,658  
 ; CURRENT FILING DATE: 1999-02-26  
 ; NUMBER OF SEQ ID NOS: 59  
 ; SOFTWARE: PatentIn ver. 2.1  
 ; SEQ ID NO 23  
 ; LENGTH: 5

Query Match 100.0%; Score 20; DB 9; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
 ||||  
 Db 1 IEGR 4

RESULT 4  
 US-09-998-831-21  
 ; Sequence 21, Application US/09998831  
 ; Patent No. US20020119153A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Phillip E. Thorpe  
 ; APPLICANT: Rolf A. Brekken  
 ; TITLE OF INVENTION: ANTI-BODY CONJUGATE COMPOSITIONS FOR SELECTIVELY  
 ; FILE REFERENCE: INHIBITING VEGF  
 ; FILE REFERENCE: 4001.002584  
 ; CURRENT APPLICATION NUMBER: US/09/998,831  
 ; CURRENT FILING DATE: 2001-11-30  
 ; PRIOR APPLICATION NUMBER: 09/561,108  
 ; PRIOR FILING DATE: 2000-04-28  
 ; NUMBER OF SEQ ID NOS: 44  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 21

Query Match 100.0%; Score 20; DB 10; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
 ||||  
 Db 1 IEGR 4

RESULT 3  
 US-09-040-518-6  
 ; Sequence 6, Application US/09040518  
 ; Patent No. US20010042255A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Karatzas, Costas N.  
 ; TITLE OF INVENTION: PRODUCTION OF BIOFILAMENTS IN TRANSGENIC  
 ; FILE REFERENCE: 08632/011001  
 ; CURRENT APPLICATION NUMBER: US/09/040,518  
 ; CURRENT FILING DATE: 1998-03-17  
 ; NUMBER OF SEQ ID NOS: 32  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 6  
 ; LENGTH: 4  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Designed to act as a recognition site for an  
 ; OTHER INFORMATION: enzyme  
 ; US-09-040-518-6

Query Match 100.0%; Score 20; DB 10; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
 ||||  
 Db 1 IEGR 4

us-10-023-888-22.rapb

Mon Apr 14 09:20:45 2003

```
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (5)
; OTHER INFORMATION: Xaa at position 5 can be any amino acid.
; OTHER INFORMATION: Description of Artificial Sequence:Factor Xa
; OTHER INFORMATION: cleavage site.
US-09-259-658-23

Query Match          100.0%; Score 20; DB 9; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4
    ||||
DB 1 IEGR 4

RESULT 7
US-09-970-308-5
; Sequence 5, Application US/09970308
; Patent No. US20020045193A1
; GENERAL INFORMATION:
; APPLICANT: BRIZZARD, BILLY L.
; APPLICANT: HERNAN, RON
; TITLE OF INVENTION: PURIFICATION OF RECOMBINANT PROTEINS FUSED TO MULTIPLE
; FILE REFERENCE: SGM 6933.2
; CURRENT APPLICATION NUMBER: US/09/970,308
; CURRENT FILING DATE: 2001-10-03
; PRIOR APPLICATION NUMBER: 09/415,000
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:synthesized
; NAME/KEY: ACT_SITE
; LOCATION: (5)
; OTHER INFORMATION: X is an amino acid except Proline or Arginine
US-09-970-308-5

Query Match          100.0%; Score 20; DB 10; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4
    ||||
DB 1 IEGR 4

RESULT 8
US-09-904-117-9
; Sequence 9, Application US/09904117
; Patent No. US20020151006A1
; GENERAL INFORMATION:
; APPLICANT: MUIR, TOM W.
; APPLICANT: COLE, PHILIP A
; APPLICANT: FRIEDMAN, JEFFREY M.
; APPLICANT: SONDHI, DOLAN
; APPLICANT: SEVERINOV, KONSTANTINE
; TITLE OF INVENTION: METHODS OF LIGATING EXPRESSED PROTEINS
; FILE REFERENCE: 600-1-214CIPB
; CURRENT APPLICATION NUMBER: US/09/904,117
; CURRENT FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/09/191,890
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-13
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/093,990
```

```
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-24
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: motif within
; OTHER INFORMATION: linker region
US-09-904-117-9

Query Match          100.0%; Score 20; DB 10; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4
    ||||
DB 1 IEGR 4

RESULT 9
US-09-899-235-29
; Sequence 29, Application US/09899235
; Patent No. US20020173620A1
; GENERAL INFORMATION:
; APPLICANT: HABERMANN, PAUL
; TITLE OF INVENTION: BIFUNCTIONAL FUSION PROTEINS FORMED FROM HIRUDIN AND
; FILE REFERENCE: 02481.1750-00
; CURRENT APPLICATION NUMBER: US/09/899,235
; CURRENT FILING DATE: 2001-07-06
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 29
; LENGTH: 6
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: peptide
US-09-899-235-29

Query Match          100.0%; Score 20; DB 9; Length 6;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4
    ||||
DB 2 IEGR 5

RESULT 10
US-10-150-262-11
; Sequence 11, Application US/10150262.
; Publication No. US20030049264A1
; GENERAL INFORMATION:
; APPLICANT: FOSTER, KEITH ALAN
; APPLICANT: DUGGAN, MICHAEL JOHN
; APPLICANT: SHONE, CLIFFORD CHARLES
; TITLE OF INVENTION: CLOSTRIDIAL TOXIN DERIVATIVES ABLE TO MODIFY
; TITLE OF INVENTION: PERIPHERAL
; TITLE OF INVENTION: SENSORY APPARENT FUNCTIONS
; FILE REFERENCE: 023223/0104
; CURRENT APPLICATION NUMBER: US/10/150,262
; CURRENT FILING DATE: 2002-05-20
; PRIOR APPLICATION NUMBER: US/09/447,356
; PRIOR FILING DATE: 1999-11-22
; PRIOR APPLICATION NUMBER: 08/945,037
; PRIOR FILING DATE: 1998-01-12
; PRIOR APPLICATION NUMBER: GB 9508204.6
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 11
```

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Organism: Spacer molecule
US-10-150-262-11
```

```
Query Match          100.0%; Score 20; DB 9; Length 7;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 IEGR 4
    ||||
Db 4 IEGR 7
```

```
RESULT 11
US-10-066-209-6
; Sequence 6, Application US/10066209
; Patent No. US20020115110A1
; GENERAL INFORMATION:
; APPLICANT: Brigham-Burke, Michael R.
; APPLICANT: Young, Peter R.
; TITLE OF INVENTION: A METHOD OF IDENTIFYING AGONIST AND
; FILE REFERENCE: ANTAGONISTS FOR TUMOR NECROSIS RELATED RECEPTORS TR1 AND TR2
; CURRENT APPLICATION NUMBER: US/10/066,209
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: 09/072,993
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/055,513
; PRIOR FILING DATE: 1997-08-13
; PRIOR APPLICATION NUMBER: 60/056,980
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/057,550
; PRIOR FILING DATE: 1997-08-29
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 8
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-10-066-209-6
```

```
Query Match          100.0%; Score 20; DB 12; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 IEGR 4
    ||||
Db 3 IEGR 6
```

```
RESULT 12
US-10-066-209-8
; Sequence 8, Application US/10066209
; Patent No. US20020115110A1
; GENERAL INFORMATION:
; APPLICANT: Brigham-Burke, Michael R.
; APPLICANT: Young, Peter R.
; TITLE OF INVENTION: A METHOD OF IDENTIFYING AGONIST AND
; FILE REFERENCE: ANTAGONISTS FOR TUMOR NECROSIS RELATED RECEPTORS TR1 AND TR2
; CURRENT APPLICATION NUMBER: US/10/066,209
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: 09/072,993
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/055,513
; PRIOR FILING DATE: 1997-08-13
; PRIOR APPLICATION NUMBER: 60/056,980
; PRIOR FILING DATE: 1997-08-26
```

```
; PRIOR APPLICATION NUMBER: 60/057,550
; PRIOR FILING DATE: 1997-08-29
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 8
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-10-066-209-8
```

```
Query Match          100.0%; Score 20; DB 12; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 IEGR 4
    ||||
Db 3 IEGR 6
```

```
RESULT 13
US-09-809-517A-2
; Sequence 2, Application US/09809517A
; Patent No. US20020034733A1
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. US20020034733A1 methods for displaying (poly)peptides/pr
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; CURRENT FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 10
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-2
```

```
Query Match          100.0%; Score 20; DB 10; Length 10;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 IEGR 4
    ||||
Db 1 IEGR 4
```

```
RESULT 14
US-10-092-908-30
; Sequence 30, Application US/10092908
; Publication No. US20030040015A1
; GENERAL INFORMATION:
; APPLICANT: Kim, Kwang-Soo
; APPLICANT: Robertson, David
; TITLE OF INVENTION: Methods and Reagents for Identifying
; TITLE OF INVENTION: Compounds and Mutations That Modulate Dopamine
; FILE REFERENCE: Beta-Hydroxylase Activity
; FILE REFERENCE: 04843/097002
; CURRENT APPLICATION NUMBER: US/10/092,908
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 60/274,095
; PRIOR FILING DATE: 2001-03-07
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 11
```

; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-092-908-30

Query Match 100.0%; Score 20; DB 9; Length 11;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
|  
|  
|  
|  
Db 1 IEGR 4

RESULT 15

US-10-092-908-46  
; Sequence 45, Application US/10092908  
; Publication No. US20030040015A1  
; GENERAL INFORMATION:  
; APPLICANT: Kim, Kwang-Hyung  
; APPLICANT: Robertson, David  
; TITLE OF INVENTION: Methods and Reagents for Identifying  
; TITLE OF INVENTION: Compounds and Mutations That Modulate Dopamine  
; TITLE OF INVENTION: Beta-Hydroxylase Activity  
; FILE REFERENCE: 04843/097002  
; CURRENT APPLICATION NUMBER: US/10/092,908  
; PRIOR FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: US 60/274,095  
; PRIOR FILING DATE: 2001-03-07  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 46  
; LENGTH: 11  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: 6  
; OTHER INFORMATION: Xaa at position 6 can be Aspartic Acid, Glutamic  
; OTHER INFORMATION: Acid or can be absent.  
US-10-092-908-46

Query Match 100.0%; Score 20; DB 9; Length 11;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
|  
|  
|  
|  
Db 1 IEGR 4

Search completed: April 13, 2003, 03:31:01  
Job time : 1.11305 secs





GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2003, 03:19:41 ; Search time 0.844555 Seconds  
(without alignments)  
3053.598 Million cell updates/sec

Title: US-10-023-888-22  
Perfect score: 20  
Sequence: 1 IEGR 4

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues  
Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_AA\_Main.\*

- 1: /cgn2\_6/ptodata/1/paa/PCTUS\_COMB.pcp.\*
- 2: /cgn2\_6/ptodata/1/paa/US06\_COMB.pcp.\*
- 3: /cgn2\_6/ptodata/1/paa/US07\_COMB.pcp.\*
- 4: /cgn2\_6/ptodata/1/paa/US08\_COMB.pcp.\*
- 5: /cgn2\_6/ptodata/1/paa/US081\_COMB.pcp.\*
- 6: /cgn2\_6/ptodata/1/paa/US082\_COMB.pcp.\*
- 7: /cgn2\_6/ptodata/1/paa/US083\_COMB.pcp.\*
- 8: /cgn2\_6/ptodata/1/paa/US084\_COMB.pcp.\*
- 9: /cgn2\_6/ptodata/1/paa/US085\_COMB.pcp.\*
- 10: /cgn2\_6/ptodata/1/paa/US086\_COMB.pcp.\*
- 11: /cgn2\_6/ptodata/1/paa/US087\_COMB.pcp.\*
- 12: /cgn2\_6/ptodata/1/paa/US088\_COMB.pcp.\*
- 13: /cgn2\_6/ptodata/1/paa/US089\_COMB.pcp.\*
- 14: /cgn2\_6/ptodata/1/paa/US090\_COMB.pcp.\*
- 15: /cgn2\_6/ptodata/1/paa/US091\_COMB.pcp.\*
- 16: /cgn2\_6/ptodata/1/paa/US092\_COMB.pcp.\*
- 17: /cgn2\_6/ptodata/1/paa/US093\_COMB.pcp.\*
- 18: /cgn2\_6/ptodata/1/paa/US094\_COMB.pcp.\*
- 19: /cgn2\_6/ptodata/1/paa/US095\_COMB.pcp.\*
- 20: /cgn2\_6/ptodata/1/paa/US096\_COMB.pcp.\*
- 21: /cgn2\_6/ptodata/1/paa/US097\_COMB.pcp.\*
- 22: /cgn2\_6/ptodata/1/paa/US098\_COMB.pcp.\*
- 23: /cgn2\_6/ptodata/1/paa/US099\_COMB.pcp.\*
- 24: /cgn2\_6/ptodata/1/paa/US100\_COMB.pcp.\*
- 25: /cgn2\_6/ptodata/1/paa/US101\_COMB.pcp.\*
- 26: /cgn2\_6/ptodata/1/paa/US102\_COMB.pcp.\*
- 27: /cgn2\_6/ptodata/1/paa/US60\_COMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	20	100.0	4	1	PCT-US02-18014-5
2	20	100.0	4	1	PCT-US02-23375-7
3	20	100.0	4	1	PCT-US02-24119-19
4	20	100.0	4	1	PCT-US02-30093-57
5	20	100.0	4	1	PCT-US94-02388-4
6	20	100.0	4	1	PCT-US96-16032-23

7	20	100.0	4	1	PCT-US97-01652-18	Sequence 18, Appl
8	20	100.0	4	1	PCT-US98-11189-13	Sequence 13, Appl
9	20	100.0	4	1	PCT-US99-01188-26	Sequence 26, Appl
10	20	100.0	4	1	PCT-US99-09258A-13	Sequence 13, Appl
11	20	100.0	4	1	PCT-US99-22756-8	Sequence 8, Appl
12	20	100.0	4	3	US-07-736-847-7	Sequence 7, Appl
13	20	100.0	4	3	US-07-920-449-2	Sequence 2, Appl
14	20	100.0	4	3	US-07-944-328-2	Sequence 2, Appl
15	20	100.0	4	3	US-07-944-328-4	Sequence 2, Appl
16	20	100.0	4	4	US-08-014-096-1	Sequence 4, Appl
17	20	100.0	4	4	US-08-040-743B-6	Sequence 1, Appl
18	20	100.0	4	4	US-08-048-977-1	Sequence 6, Appl
19	20	100.0	4	4	US-08-078-153-15	Sequence 1, Appl
20	20	100.0	4	4	US-08-082-943-1	Sequence 15, Appl
21	20	100.0	4	6	US-08-235-515-19	Sequence 1, Appl
22	20	100.0	4	6	US-08-252-820A-15	Sequence 19, Appl
23	20	100.0	4	6	US-08-252-820B-15	Sequence 15, Appl
24	20	100.0	4	6	US-08-278-774-2	Sequence 15, Appl
25	20	100.0	4	7	US-08-302-892-1	Sequence 2, Appl
26	20	100.0	4	7	US-08-347-893-8	Sequence 1, Appl
27	20	100.0	4	7	US-08-350-528-16	Sequence 8, Appl
28	20	100.0	4	7	US-08-350-530-4	Sequence 16, Appl
29	20	100.0	4	7	US-08-350-530A-4	Sequence 4, Appl
30	20	100.0	4	8	US-08-438-852-2	Sequence 4, Appl
31	20	100.0	4	8	US-08-445-427-6	Sequence 2, Appl
32	20	100.0	4	8	US-08-446-650-3	Sequence 6, Appl
33	20	100.0	4	8	US-08-453-485E-14	Sequence 3, Appl
34	20	100.0	4	8	US-08-468-131-8	Sequence 14, Appl
35	20	100.0	4	8	US-08-471-052-19	Sequence 8, Appl
36	20	100.0	4	8	US-08-471-927-19	Sequence 19, Appl
37	20	100.0	4	8	US-08-477-512A-32	Sequence 19, Appl
38	20	100.0	4	8	US-08-486-817A-32	Sequence 32, Appl
39	20	100.0	4	9	US-08-540-397-38	Sequence 32, Appl
40	20	100.0	4	9	US-08-595-868A-18	Sequence 38, Appl
41	20	100.0	4	10	US-08-626-288-3	Sequence 18, Appl
42	20	100.0	4	10	US-08-643-731-88	Sequence 3, Appl
43	20	100.0	4	10	US-08-680-004-8	Sequence 88, Appl
44	20	100.0	4	12	US-08-865-398-22	Sequence 8, Appl
45	20	100.0	4	13	US-08-909-601-33	Sequence 22, Appl

## ALIGNMENTS

RESULT 1  
PCT-US02-18014-5  
; Sequence 5, Application PC/TUS0218014  
; GENERAL INFORMATION:  
; APPLICANT: University of Utah Research Foundation  
; APPLICANT: Rechsteiner, Martin  
; APPLICANT: Pratt, Gregory  
; APPLICANT: Li, Jun  
; TITLE OF INVENTION: METHODS FOR IDENTIFYING COMPOUNDS WITH  
; TITLE OF INVENTION: THERAPEUTIC POTENTIAL FOR TREATMENT OF CENTRAL  
; TITLE OF INVENTION: NEURODEGENERATIVE DISEASES RESULTING FROM ABNORMAL PROTEIN  
; FILE REFERENCE: 21101.0013p1  
; CURRENT APPLICATION NUMBER: PCT/US02/18014  
; PRIOR FILING DATE: 2002-06-07  
; PRIOR APPLICATION NUMBER: 60/297,332  
; PRIOR FILING DATE: 2001-06-11  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence./note -  
; OTHER INFORMATION: synthetic construct  
PCT-US02-18014-5

Query Match 100.0% Score 20; DB 1; Length 4;

Best Local Similarity 100.0%; Pred. No. 4.2e+06; Indels 0; Gaps 0; Mismatches 0;

Matches 4; Conservative 0;

QY 1 IEGR 4  
| | | |

Db 1 IEGR 4

RESULT 2

PCT-US02-23375-7

; Sequence 7, Application PC/TUS0223375

; GENERAL INFORMATION:

; APPLICANT: Hagen, Frederick S.

; APPLICANT: Woodbury, Richard G.

; APPLICANT: Oort, Pieter J.

; APPLICANT: IcoGen Corporation

; TITLE OF INVENTION: METHODS FOR THE IDENTIFICATION OF PEPTIDYL COMPOUNDS

; TITLE OF INVENTION: INTERACTING WITH EXTRACELLULAR TARGET MOLECULES

; FILE REFERENCE: 17881-7-1PC

; CURRENT APPLICATION NUMBER: PCT/US02/23375

; CURRENT FILING DATE: 2002-07-19

; PRIOR APPLICATION NUMBER: 60/306,924

; PRIOR FILING DATE: 2001-07-19

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 7

; LENGTH: 4

; TYPE: PRT

; ORGANISM: Factor Xa recognition sequence

PCT-US02-23375-7

Query Match 100.0%; Score 20; DB 1; Length 4;

Best Local Similarity 100.0%; Pred. No. 4.2e+06; Indels 0; Gaps 0; Mismatches 0;

Matches 4; Conservative 0;

QY 1 IEGR 4  
| | | |

Db 1 IEGR 4

RESULT 3

PCT-US02-24119-19

; Sequence 19, Application PC/TUS0224119

; GENERAL INFORMATION:

; APPLICANT: Haugwitz, Michael

; TITLE OF INVENTION: Method of Detecting Protease Activity in

; FILE REFERENCE: CLON-082WO

; CURRENT APPLICATION NUMBER: PCT/US02/24119

; CURRENT FILING DATE: 2002-07-30

; PRIOR APPLICATION NUMBER: 60/309,312

; PRIOR FILING DATE: 2001-07-31

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 19

; LENGTH: 4

; TYPE: PRT

; ORGANISM: human

PCT-US02-24119-19

Query Match 100.0%; Score 20; DB 1; Length 4;

Best Local Similarity 100.0%; Pred. No. 4.2e+06; Indels 0; Gaps 0; Mismatches 0;

Matches 4; Conservative 0;

QY 1 IEGR 4  
| | | |

Db 1 IEGR 4

RESULT 4

PCT-US02-30093-57

; Sequence 57, Application PC/TUS0230093

; GENERAL INFORMATION:

; APPLICANT: University of Rochester

; APPLICANT: Puzas, J. Edward

; TITLE OF INVENTION: COMPOSITIONS AND METHODS INVOLVED IN

; TITLE OF INVENTION: BONE GROWTH

; FILE REFERENCE: 21108.0004P1

; CURRENT APPLICATION NUMBER: PCT/US02/30093

; CURRENT FILING DATE: 2002-09-20

; PRIOR APPLICATION NUMBER: 60/323,987

; PRIOR FILING DATE: 2001-09-20

; NUMBER OF SEQ ID NOS: 78

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 57

; LENGTH: 4

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence:/note =

; OTHER INFORMATION: Synthetic Construct

PCT-US02-30093-57

Query Match 100.0%; Score 20; DB 1; Length 4;

Best Local Similarity 100.0%; Pred. No. 4.2e+06; Indels 0; Gaps 0; Mismatches 0;

Matches 4; Conservative 0;

QY 1 IEGR 4  
| | | |

Db 1 IEGR 4

RESULT 5

PCT-US94-02388-4

; Sequence 4, Application PC/TUS9402388

; GENERAL INFORMATION:

; APPLICANT: TURNER, George J.

; APPLICANT: BETLACH, Mary C.

; TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS POLYPEPTIDES

; TITLE OF INVENTION: IN HALOBACTERIA

; NUMBER OF SEQUENCES: 15

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Robert Berliner

; STREET: 201 North Figueroa Street

; CITY: Los Angeles

; STATE: California

; COUNTRY: USA

; ZIP: 90012

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US94/02388

; FILING DATE:

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Berliner, Robert

; REGISTRATION NUMBER: 20,121

; REFERENCE/DOCKET NUMBER: 5555-206-PCT

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (213) 977-1001

; TELEFAX: (213) 977-1003

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 4 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

PCT-US94-02388-4

Query Match 100.0%; Score 20; DB 1; Length 4;

Best Local Similarity 100.0%; Pred. No. 4.2e+06; Indels 0; Gaps 0; Mismatches 0;

Matches 4; Conservative 0;

QY 1 IEGR 4  
||||  
Db 1 IEGR 4

## RESULT 6

PCT-US96-16032-23  
; Sequence 23, Application PC/TUS9616032  
; GENERAL INFORMATION:  
; APPLICANT: PENCE  
; TITLE OF INVENTION: Coiled-Coil Heterodimer Methods and  
; TITLE OF INVENTION: Compositions for the Detection and Purification  
; TITLE OF INVENTION: of Expressed Polypeptides  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dehlinger & Associates  
; STREET: 350 Cambridge Avenue, Suite 250  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US96/16032  
; FILING DATE: 04-OCT-1996  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/540,397  
; FILING DATE: 06-OCT-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sholtz, Charles K.  
; REGISTRATION NUMBER: 38,615  
; REFERENCE/DOCKET NUMBER: 8900-0109.41  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 324-0880  
; TELEFAX: (415) 324-0960  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: Factor Xa cleavage site  
PCT-US96-16032-23

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 4.2e+06;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
||||  
Db 1 IEGR 4

## RESULT 7

PCT-US97-01652-18  
; Sequence 18, Application PC/TUS9701652  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: RECOMBINANT PREPARATION OF CALCITONIN FRAGMENTS AND US  
; NUMBER OF SEQUENCES: 50  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Merchant & Gould  
; STREET: 3100 Norwest Center, 90 S. 7th Street  
; CITY: Minneapolis  
; STATE: Minnesota  
; COUNTRY: United States of America

ZIP: 55402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSEQ Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US97/01652  
; FILING DATE: 04-FEB-1997  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/595,868  
; FILING DATE: 06-FEB-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CARTER, Charles G. / BRUESS, Steven C.  
; REGISTRATION NUMBER: 35,093 / 34,130  
; REFERENCE/DOCKET NUMBER: 8648.59W001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 612/371-5278 / 612/336-4711  
; TELEFAX: 612/336-4751  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: Internal  
; ORIGINAL SOURCE:  
PCT-US97-01652-18

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 4.2e+06;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
||||  
Db 1 IEGR 4

## RESULT 8

PCT-US98-11189-13  
; Sequence 13, Application PC/TUS9811189A  
; GENERAL INFORMATION:  
; APPLICANT: Marcel E. Nimni  
; APPLICANT: Frederick L. Hall  
; APPLICANT: Lingtao Wu  
; APPLICANT: Bo Han  
; APPLICANT: Edwin Shors  
; TITLE OF INVENTION: BONE MORPHOGENETIC PROTEINS AND THEIR  
; TITLE OF INVENTION: USE IN BONE GROWTH  
; FILE REFERENCE: 17972-11  
; CURRENT APPLICATION NUMBER: PCT/US98/11189A  
; CURRENT FILING DATE: 1998-06-02  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 13  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Human  
PCT-US98-11189-13

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 4.2e+06;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
||||  
Db 1 IEGR 4

```
RESULT 9
PCT-US99-01188-26
; Sequence 26, Application PC/TUS9901188
; GENERAL INFORMATION:
; APPLICANT: Wittup et al.
; TITLE OF INVENTION: Yeast Cell Surface Display of Proteins and Uses
; FILE OF INVENTION: Thereof
; FILE REFERENCE: D6061PCT
; CURRENT APPLICATION NUMBER: PCT/US99/01188
; CURRENT FILING DATE: 1999-01-20
; EARLIER APPLICATION NUMBER: 09/009,388
; EARLIER FILING DATE: 1998-01-20
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: WORD 6.0.1 for Macintosh
; SEQ ID NO 26
; LENGTH: 4
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Epitope tag
PCT-US99-01188-26

Query Match      100.0%; Score 20; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 4.2e+06;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4
   ||||
Db 1 IEGR 4

RESULT 10
PCT-US99-09256A-13
; Sequence 13, Application PC/TUS9909256A
; GENERAL INFORMATION:
; APPLICANT: Maty Aval-Herskovitz et al.
; TITLE OF INVENTION: GENETICALLY MODIFIED CELLS AND METHODS
; FILE OF INVENTION: FOR EXPRESSING RECOMBINANT HEPARANASE
; TITLE OF INVENTION: AND METHODS OF PURIFYING SAME
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mark M. Friedman c/o Anthony Castorina
; STREET: 2001 Jefferson Davis Highway, Suite 207
; CITY: Arlington
; STATE: Virginia
; COUNTRY: United States of America
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
; COMPUTER: Twinhead* Slimnote-890TX
; OPERATING SYSTEM: MS DOS version 6.2,
; OPERATING SYSTEM: Windows version 3.11
; SOFTWARE: Word for Windows version 2.0 converted to
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US99/09256A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/071,618
; FILING DATE: May 1, 1998
; APPLICATION NUMBER: 09/071,739
; FILING DATE: May 1, 1998
; APPLICATION NUMBER: 08/922,180
; FILING DATE: September 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Friedman, Mark M.
; REGISTRATION NUMBER: 33,883
; REFERENCE/DOCKET NUMBER: 910/16
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 972-3-562553
; TELEFAX: 972-3-562554
```

```
TELEX:
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
PCT-US99-09256A-13

Query Match      100.0%; Score 20; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 4.2e+06;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4
   ||||
Db 1 IEGR 4

RESULT 11
PCT-US99-22756-8
; Sequence 8, Application PC/TUS9922756
; GENERAL INFORMATION:
; APPLICANT: Fronticelli, Clara
; TITLE OF INVENTION: POLYMERIC HEMOGLOBIN MUTANTS
; FILE REFERENCE: 6056-279 PC
; CURRENT APPLICATION NUMBER: PCT/US99/22756
; CURRENT FILING DATE: 1999-09-30
; EARLIER APPLICATION NUMBER: 60/102,640
; EARLIER FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 4
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Factor Xa
; OTHER INFORMATION: recognition sequence
PCT-US99-22756-8

Query Match      100.0%; Score 20; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 4.2e+06;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4
   ||||
Db 1 IEGR 4

RESULT 12
US-07-736-847-7
; Sequence 7, Application US/07736847
; GENERAL INFORMATION:
; APPLICANT: Blonar, Michael A.
; APPLICANT: Kaelin, William G.
; TITLE OF INVENTION: Plasmids for the Rapid Preparation of
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/736,847
; FILING DATE: 19910729
```

/ CLASSIFICATION: 435  
/ ATTORNEY/AGENT INFORMATION:  
/ NAME: Brook, David E.  
/ REGISTRATION NUMBER: 22,592  
/ REFERENCE/DOCKET NUMBER: DFCI-208  
/ TELECOMMUNICATION INFORMATION:  
/ TELEPHONE: (617) 861-6240  
/ TELEFAX: (617) 861-9540  
/ TELEX: 951794  
/ INFORMATION FOR SEQ ID NO: 7:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 4 amino acids  
/ TYPE: AMINO ACID  
/ TOPOLOGY: linear  
/ MOLECULE TYPE: peptide  
/ US-07-736-847-7

Query Match 100.0%; Score 20; DB 3; Length 4;  
Best Local Similarity 100.0%; Pred. No. 4.2e+06;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
Db 1 IEGR 4

RESULT 13  
US-07-920-449-2  
/ SEQUENCE 2, Application US/07920449  
/ GENERAL INFORMATION:  
/ APPLICANT: LAU, Peter  
/ APPLICANT: RIOUX, Clement  
/ TITLE OF INVENTION: LIPOPROTEIN SIGNAL PEPTIDE FUSED TO  
/ TITLE OF INVENTION: ANTIGENIC POLYPEPTIDES  
/ NUMBER OF SEQUENCES: 13  
/ CORRESPONDENCE ADDRESS:  
/ ADDRESSEE: Foley & Lardner  
/ STREET: 3000 K Street, N.W., Suite 500  
/ CITY: Washington, D.C.  
/ COUNTRY: USA  
/ ZIP: 20007-5109

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/920,449  
FILING DATE: 19920829  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/CA90/00460  
FILING DATE: 27-DEC-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: SAXE, Bernhard D.  
REGISTRATION NUMBER: 28,665  
REFERENCE/DOCKET NUMBER: 32931/106 SMAB  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear

US-07-920-449-2

Query Match 100.0%; Score 20; DB 3; Length 4;  
Best Local Similarity 100.0%; Pred. No. 4.2e+06;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4

Db 1 IEGR 4

RESULT 14  
US-07-944-328-2  
/ SEQUENCE 2, Application US/07944328  
/ GENERAL INFORMATION:  
/ APPLICANT: Sakamoto, Takeshi  
/ APPLICANT: Takamoto, Kazunori  
/ APPLICANT: Senda, Toshiya  
/ APPLICANT: Harada, Yoshinori  
/ TITLE OF INVENTION: FACTOR Xa LINKER DNA  
/ NUMBER OF SEQUENCES: 4  
/ CORRESPONDENCE ADDRESS:  
/ ADDRESSEE: Antonelli, Terry, Stout & Kraus  
/ STREET: 1919 Pennsylvania, Ave., N.W.; Suite 600  
/ CITY: Washington  
/ STATE: D.C.  
/ COUNTRY: U.S.A.  
/ ZIP: 20006

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/944,328  
FILING DATE: 19920914  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP-A-3-234430  
FILING DATE: September 13, 1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Terry, David T.  
REGISTRATION NUMBER: 20,178  
REFERENCE/DOCKET NUMBER: 500.31492XOO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-828-0300  
TELEFAX: 202-828-0380  
TELEX: 440280

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-07-944-328-2

Query Match 100.0%; Score 20; DB 3; Length 4;  
Best Local Similarity 100.0%; Pred. No. 4.2e+06;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
Db 1 IEGR 4

RESULT 15  
US-07-944-328-4  
/ SEQUENCE 4, Application US/07944328  
/ GENERAL INFORMATION:  
/ APPLICANT: Sakamoto, Takeshi  
/ APPLICANT: Takamoto, Kazunori  
/ APPLICANT: Senda, Toshiya  
/ APPLICANT: Harada, Yoshinori  
/ TITLE OF INVENTION: FACTOR Xa LINKER DNA  
/ NUMBER OF SEQUENCES: 4  
/ CORRESPONDENCE ADDRESS:  
/ ADDRESSEE: Antonelli, Terry, Stout & Kraus  
/ STREET: 1919 Pennsylvania, Ave., N.W.; Suite 600  
/ CITY: Washington  
/ STATE: D.C.

```

; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/944,328
; FILING DATE: 19920914
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP-A-3-234430
; FILING DATE: September 13, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Terry, David T.
; REGISTRATION NUMBER: 20,178
; REFERENCE/DOCKET NUMBER: 500.31492X00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-828-0300
; TELEFAX: 202-828-0380
; TELEX: 440280
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-944-328-4

```

```

Query Match      100.0%; Score 20; DB 3; Length 4;
Best Local Similarity 100.0%; Pred. No. 4.2e+06;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 IEGR 4
Db 1 IEGR 4

```

```

Search completed: April 13, 2003, 03:28:06
Job time : 1.84456 secs

```

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2003, 03:21:45 ; Search time 0.202826 Seconds  
(without alignments)  
3053.886 Million cell updates/sec

Title: US-10-023-888-22  
Perfect score: 20  
Sequence: 1 IEGR 4

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 696924 seqs, 154852082 residues

Total number of hits satisfying chosen parameters: 696924

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_AA\_New:\*  
1: /cgn2\_6/ptodata/2/paa/PCT\_NEW\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/paa/US06\_NEW\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/paa/US07\_NEW\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/paa/US08\_NEW\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pep:\*  
7: /cgn2\_6/ptodata/2/paa/US60\_NEW\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query %			DB	ID	Description
	Score	Match	Length			
1	20	100.0	4	1	PCT-US02-33563-8	Sequence 8, Appl
2	20	100.0	4	1	PCT-US02-29264-35	Sequence 35, Appl
3	20	100.0	4	1	PCT-US03-04213-6	Sequence 6, Appl
4	20	100.0	4	1	PCT-US02-37624A-22	Sequence 22, Appl
5	20	100.0	4	5	US-09-516-055-17	Sequence 17, Appl
6	20	100.0	4	5	US-09-787-216A-8	Sequence 8, Appl
7	20	100.0	4	5	US-09-445-576A-104	Sequence 104, App
8	20	100.0	4	5	US-09-640-198D-28	Sequence 28, Appl
9	20	100.0	4	5	US-09-858-332A-6	Sequence 6, Appl
10	20	100.0	4	5	US-09-858-332B-6	Sequence 6, Appl
11	20	100.0	4	5	US-09-194-223B-150	Sequence 150, App
12	20	100.0	4	6	US-10-128-587A-70	Sequence 70, Appl
13	20	100.0	4	6	US-10-274-638-8	Sequence 8, Appl
14	20	100.0	4	6	US-10-087-775-15	Sequence 15, Appl
15	20	100.0	4	6	US-10-328-813-10	Sequence 10, Appl
16	20	100.0	4	6	US-10-158-742A-18	Sequence 18, Appl
17	20	100.0	4	6	US-10-311-406-5	Sequence 5, Appl
18	20	100.0	4	6	US-10-343-977-6	Sequence 6, Appl
19	20	100.0	4	6	US-10-365-095-6	Sequence 6, Appl
20	20	100.0	4	6	US-10-369-779-45	Sequence 45, Appl
21	20	100.0	4	6	US-10-373-561-21	Sequence 21, Appl
22	20	100.0	4	6	US-10-375-913-16	Sequence 16, Appl
23	20	100.0	5	4	US-08-981-087B-8	Sequence 8, Appl
24	20	100.0	5	5	US-09-699-314-1	Sequence 1, Appl
25	20	100.0	5	6	US-10-128-587A-66	Sequence 66, Appl
26	20	100.0	5	6	US-10-128-587A-69	Sequence 69, Appl

27 20 100.0 5 6 US-10-128-578B-64 Sequence 64, Appl  
28 20 100.0 6 5 US-09-445-576A-10 Sequence 10, Appl  
29 20 100.0 7 5 US-09-194-223B-71 Sequence 71, Appl  
30 20 100.0 8 6 US-10-360-053-29 Sequence 29, Appl  
31 20 100.0 9 5 US-09-445-576A-77 Sequence 77, Appl  
32 20 100.0 9 5 US-09-445-576A-86 Sequence 86, Appl  
33 20 100.0 9 6 US-10-363-205-106 Sequence 106, Appl  
34 20 100.0 10 5 US-09-445-576A-82 Sequence 82, Appl  
35 20 100.0 10 6 US-10-130-973A-18 Sequence 18, Appl  
36 20 100.0 13 1 PCT-US02-09017-11 Sequence 11, Appl  
37 20 100.0 13 1 PCT-US02-09017A-11 Sequence 11, Appl  
38 20 100.0 13 5 US-09-935-061-4 Sequence 4, Appl  
39 20 100.0 13 5 US-09-689-516C-11 Sequence 11, Appl  
40 20 100.0 13 5 US-09-194-223B-73 Sequence 73, Appl  
41 20 100.0 13 5 US-09-194-223B-85 Sequence 85, Appl  
42 20 100.0 14 1 PCT-US02-09017-9 Sequence 9, Appl  
43 20 100.0 14 1 PCT-US02-09017-10 Sequence 10, Appl  
44 20 100.0 14 1 PCT-US02-09017A-9 Sequence 9, Appl  
45 20 100.0 14 1 PCT-US02-09017A-10 Sequence 10, Appl

## ALIGNMENTS

RESULT 1  
PCT-US02-33563-8  
; Sequence 8, Application PC/TUS0233563  
; GENERAL INFORMATION:  
; APPLICANT: ZymoGenetics, Inc.  
; TITLE OF INVENTION: DIMERIZED GROWTH FACTOR AND MATERIALS  
; FILE REFERENCE: 01-30FC  
; CURRENT APPLICATION NUMBER: PCT/US02/33563  
; CURRENT FILING DATE: 2002-10-18  
; PRIOR APPLICATION NUMBER: 60/346,117  
; PRIOR FILING DATE: 2001-10-19  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: polypeptide, factor Xa cleavage site  
PCT-US02-33563-8

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 5.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
||||  
Db 1 IEGR 4

RESULT 2  
PCT-US02-29264-35  
; Sequence 35, Application PC/TUS0229264  
; GENERAL INFORMATION:  
; APPLICANT: Invitrogen Corporation  
; TITLE OF INVENTION: DNA Polymerases and Mutants Thereof  
; FILE REFERENCE: 0942.536PC01  
; CURRENT APPLICATION NUMBER: PCT/US02/29264  
; CURRENT FILING DATE: 2002-09-16  
; PRIOR APPLICATION NUMBER: US 60/318,903  
; PRIOR FILING DATE: 2001-09-14  
; NUMBER OF SEQ ID NOS: 3639  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 35  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:



; OTHER INFORMATION: Factor Xa cleavage site  
PCT-US02-29264-35

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 5.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
|  
|  
|  
|  
Db 1 IEGR 4

## RESULT 3

PCT-US03-04213-6  
; Sequence 6, Application PC/TUS0304213  
; GENERAL INFORMATION:  
; APPLICANT: ZymoGenetics, Inc.  
; TITLE OF INVENTION: MATERIALS AND METHODS FOR PREPARING DIMERIC GROWTH FACTORS  
; FILE REFERENCE: 01-33PC  
; CURRENT APPLICATION NUMBER: PCT/US03/04213  
; CURRENT FILING DATE: 2003-02-11  
; PRIOR APPLICATION NUMBER: 60/355,882  
; PRIOR FILING DATE: 2002-02-11  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: factor Xa cleavage site  
PCT-US03-04213-6

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 5.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
|  
|  
|  
|  
Db 1 IEGR 4

## RESULT 4

PCT-US02-37624A-22  
; Sequence 22, Application PC/TUS0237624A  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: SOLUBLE GLCNAC PHOSPHOTRANSFERASE  
; FILE REFERENCE: 203515US77  
; CURRENT APPLICATION NUMBER: PCT/US02/37624A  
; CURRENT FILING DATE: 2003-02-27  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 22  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: synthetic peptide  
PCT-US02-37624A-22

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 5.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
|  
|  
|  
|  
Db 1 IEGR 4

## RESULT 5

US-09-516-055-17  
; Sequence 17, Application US/09516055

; GENERAL INFORMATION:  
; APPLICANT: Karatzas, Costas A.  
; APPLICANT: Huang, Yue-Jin  
; TITLE OF INVENTION: TRANSGENIC ANIMALS THAT PRODUCE ALTERED  
; TITLE OF INVENTION: WOOL  
; FILE REFERENCE: 06632/016002  
; CURRENT APPLICATION NUMBER: US/09/516,055  
; CURRENT FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: US 60/149,354  
; PRIOR FILING DATE: 1999-08-18  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic  
US-09-516-055-17

Query Match 100.0%; Score 20; DB 5; Length 4;  
Best Local Similarity 100.0%; Pred. No. 5.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
|  
|  
|  
|  
Db 1 IEGR 4

## RESULT 6

US-09-787-216A-8  
; Sequence 8, Application US/09787216A  
; GENERAL INFORMATION:  
; APPLICANT: Fronticelli, Clara  
; TITLE OF INVENTION: POLYMERIC HEMOGLOBIN MUTANTS  
; FILE REFERENCE: 6056-279 PC  
; CURRENT APPLICATION NUMBER: US/09/787,216A  
; CURRENT FILING DATE: 2002-09-20  
; PRIOR APPLICATION NUMBER: PCT/US99/22756  
; PRIOR FILING DATE: 1999-09-30  
; PRIOR APPLICATION NUMBER: 60/102,640  
; PRIOR FILING DATE: 1998-10-01  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Factor Xa  
; OTHER INFORMATION: recognition sequence  
US-09-787-216A-8

Query Match 100.0%; Score 20; DB 5; Length 4;  
Best Local Similarity 100.0%; Pred. No. 5.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
|  
|  
|  
|  
Db 1 IEGR 4

## RESULT 7

US-09-445-576A-104  
; Sequence 104, Application US/09445576A  
; GENERAL INFORMATION:  
; APPLICANT: Borean Pharma A/S  
; TITLE OF INVENTION: Trimerising module  
; FILE REFERENCE: 62032.000004  
; CURRENT APPLICATION NUMBER: US/09/445,576A  
; CURRENT FILING DATE: 2000-07-17  
; NUMBER OF SEQ ID NOS: 104  
; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 104  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Fxa protease site  
US-09-445-576A-104

Query Match 100.0%; Score 20; DB 5; Length 4;  
Best Local Similarity 100.0%; Pred. No. 5.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
|  
|  
|  
|  
Db 1 IEGR 4

RESULT 8  
US-09-640-198D-28  
; Sequence 28, Application US/09640198D  
; GENERAL INFORMATION:  
; APPLICANT: Russell, Stephen  
; APPLICANT: Kay Whye, Peng  
; TITLE OF INVENTION: System for Monitoring the Location of  
; TITLE OF INVENTION: Transgenes  
; FILE REFERENCE: 07039-295001  
; CURRENT APPLICATION NUMBER: US/09/640.198D  
; PRIOR FILING DATE: 2000-08-16  
; PRIOR APPLICATION NUMBER: US 60/149,168  
; PRIOR FILING DATE: 1999-08-17  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 28  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Factor Xa Cleavage Site  
US-09-640-198D-28

Query Match 100.0%; Score 20; DB 5; Length 4;  
Best Local Similarity 100.0%; Pred. No. 5.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
|  
|  
|  
|  
Db 1 IEGR 4

RESULT 9  
US-09-858-332A-6  
; Sequence 6, Application US/09858332A  
; GENERAL INFORMATION:  
; APPLICANT: Tchaga, Grigory S.  
; APPLICANT: Jokhadze, George  
; TITLE OF INVENTION: Metal Ion Affinity Tags and Methods for  
; TITLE OF INVENTION: Using the Same  
; FILE REFERENCE: CLON-056CIP  
; CURRENT APPLICATION NUMBER: US/09/858.332A  
; PRIOR FILING DATE: 2001-05-15  
; PRIOR APPLICATION NUMBER: 09/404,017  
; PRIOR FILING DATE: 1999-09-23  
; PRIOR APPLICATION NUMBER: 60/101,867  
; PRIOR FILING DATE: 1998-09-25  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: enterokinase cleavage site  
US-09-858-332A-6

Query Match 100.0%; Score 20; DB 5; Length 4;  
Best Local Similarity 100.0%; Pred. No. 5.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
|  
|  
|  
|  
Db 1 IEGR 4

RESULT 10  
US-09-858-332B-6  
; Sequence 6, Application US/09858332B  
; GENERAL INFORMATION:  
; APPLICANT: Tchaga, Grigory S.  
; APPLICANT: Jokhadze, George  
; TITLE OF INVENTION: Metal Ion Affinity Tags and Methods for  
; TITLE OF INVENTION: Using the Same  
; FILE REFERENCE: CLON-056CIP  
; CURRENT APPLICATION NUMBER: US/09/858.332B  
; CURRENT FILING DATE: 2001-05-15  
; PRIOR APPLICATION NUMBER: 09/404,017  
; PRIOR FILING DATE: 1999-09-23  
; PRIOR APPLICATION NUMBER: 60/101,867  
; PRIOR FILING DATE: 1998-09-25  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: enterokinase cleavage site  
US-09-858-332B-6

Query Match 100.0%; Score 20; DB 5; Length 4;  
Best Local Similarity 100.0%; Pred. No. 5.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
|  
|  
|  
|  
Db 1 IEGR 4

RESULT 11  
US-09-194-223B-150  
; Sequence 150, Application US/09194223B  
; GENERAL INFORMATION:  
; APPLICANT: COSSET, FRANCOIS-LOIC  
; APPLICANT: VALSESLA, SANDRINE  
; APPLICANT: RUSSELL, STEPHEN J.  
; TITLE OF INVENTION: VIRAL PARTICLES WHICH ARE MASKED OR UNMASKED WITH  
; TITLE OF INVENTION: RESPECT TO A CELL RECEPTOR  
; FILE REFERENCE: 0508-1017  
; CURRENT APPLICATION NUMBER: US/09/194.223B  
; CURRENT FILING DATE: 1999-03-23  
; PRIOR APPLICATION NUMBER: PCT/FR97/00870  
; PRIOR FILING DATE: 1997-05-16  
; PRIOR APPLICATION NUMBER: FR 96.06234  
; PRIOR FILING DATE: 1996-05-20  
; NUMBER OF SEQ ID NOS: 150  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 150  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
; OTHER INFORMATION: peptide  
US-09-194-223B-150

Query Match 100.0%; Score 20; DB 5; Length 4;  
Best Local Similarity 100.0%; Pred. No. 5.9e+05;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4

Db 1 IEGR 4

## RESULT 12

US-10-128-587A-70  
; Sequence 70, Application US/10128587A  
; GENERAL INFORMATION:  
; APPLICANT: Innogenetics N.V.  
; TITLE OF INVENTION: Constructs and methods for expression of recombinant HCV envelope  
; FILE REFERENCE: 134 PCT  
; CURRENT APPLICATION NUMBER: US/10/128,587A  
; PRIOR FILING DATE: 2002-04-24  
; NUMBER OF SEQ ID NOS: 98  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 70  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: adaptor peptide  
US-10-128-587A-70

Query Match 100.0%; Score 20; DB 6; Length 4;  
Best Local Similarity 100.0%; Pred. No. 5.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4

Db 1 IEGR 4

## RESULT 13

US-10-274-638-8  
; Sequence 8, Application US/10274638  
; GENERAL INFORMATION:  
; APPLICANT: Moore, Margaret D.  
; APPLICANT: Fox, Brian A.  
; TITLE OF INVENTION: DIMERIZED GROWTH FACTOR AND MATERIALS  
; FILE REFERENCE: 01-30  
; CURRENT APPLICATION NUMBER: US/10/274,638  
; PRIOR FILING DATE: 2002-10-18  
; PRIOR FILING DATE: 2001-10-19  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: polypeptide, factor xa cleavage site  
US-10-274-638-8

Query Match 100.0%; Score 20; DB 6; Length 4;  
Best Local Similarity 100.0%; Pred. No. 5.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4

Db 1 IEGR 4

## RESULT 14

US-10-087-775-15  
; Sequence 15, Application US/10087775  
; GENERAL INFORMATION:  
; APPLICANT: INABA, Niro

; APPLICANT: HORI, Takeya  
; APPLICANT: ITO, Satoru  
; TITLE OF INVENTION: Method for Producing Proteins  
; FILE REFERENCE: 0760-0303P  
; CURRENT APPLICATION NUMBER: US/10/087,775  
; CURRENT FILING DATE: 2002-03-05  
; PRIOR APPLICATION NUMBER: JP 2001-60973  
; PRIOR FILING DATE: 2001-03-05  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 15  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Intervening peptide region recognized by Factor Xa.  
US-10-087-775-15

Query Match 100.0%; Score 20; DB 6; Length 4;  
Best Local Similarity 100.0%; Pred. No. 5.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4

Db 1 IEGR 4

## RESULT 15

US-10-328-813-10  
; Sequence 10, Application US/10328813  
; GENERAL INFORMATION:  
; APPLICANT: Osborne, William R.A.  
; APPLICANT: Ramesh, Nagarajan  
; TITLE OF INVENTION: Compositions and Methods for Treating Diabetes  
; FILE REFERENCE: P-UW 3264  
; CURRENT APPLICATION NUMBER: US/10/328,813  
; CURRENT FILING DATE: 2002-12-23  
; PRIOR APPLICATION NUMBER: US/09/185,852  
; PRIOR FILING DATE: 1998-11-04  
; PRIOR APPLICATION NUMBER: 60/087,660  
; PRIOR FILING DATE: 1998-06-02  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 10  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Consensus  
; OTHER INFORMATION: Sequence  
US-10-328-813-10

Query Match 100.0%; Score 20; DB 6; Length 4;  
Best Local Similarity 100.0%; Pred. No. 5.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4

Db 1 IEGR 4

Search completed: April 13, 2003, 03:30:10  
Job time : 1.20283 secs

GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2003, 03:03:50 ; Search time 0.129676 Seconds  
(without alignments)  
2965.377 Million cell updates/sec

Title: US-10-023-888-22

Perfect score: 20

Sequence: 1 IEGR 4

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_73.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	63	2 S24422	4-oxalocrotonate t
2	20	100.0	66	2 C70202	hypothetical prote
3	20	100.0	67	2 A13396	hypothetical prote
4	20	100.0	71	2 D82886	hypothetical prote
5	20	100.0	73	2 T10269	hydroxymethylgluta
6	20	100.0	73	2 H69460	conserved hypothet
7	20	100.0	74	2 S53481	FUN57 protein - ye
8	20	100.0	75	2 AC2441	hypothetical prote
9	20	100.0	75	2 AD2183	hypothetical prote
10	20	100.0	77	2 B83961	acyl-carrier prote
11	20	100.0	77	2 T49820	hypothetical prote
12	20	100.0	80	2 AE1946	hypothetical prote
13	20	100.0	82	2 T01133	hypothetical prote
14	20	100.0	86	2 D71635	hypothetical prote
15	20	100.0	86	2 C97845	ribosomal protein
16	20	100.0	86	2 F85354	50S ribosomal prot
17	20	100.0	88	2 A84568	small nuclear ribo
18	20	100.0	92	2 B90356	probable small nuc
19	20	100.0	94	2 A83523	DNA-directed RNA p
20	20	100.0	101	2 E72458	hypothetical prote
21	20	100.0	102	2 T10099	hypothetical prote
22	20	100.0	102	2 B97736	probable hydroxyme
23	20	100.0	105	2 H87543	hypothetical prote
24	20	100.0	106	2 G96534	conserved hypothet
25	20	100.0	106	2 C90261	hypothetical prote
26	20	100.0	107	2 B97370	hypothetical prote
27	20	100.0	107	2 AB2588	hypothetical prote
28	20	100.0	109	2 JC2440	conserved hypothet
29	20	100.0	109	2 H71142	biphenyl dioxygena
					hypothetical prote

30 20 100.0 109 2 G70317 conserved hypothet  
31 20 100.0 111 2 G75009 hypothetical prote  
32 20 100.0 111 2 AB1383 hypothetical prote  
33 20 100.0 113 2 C69045 hypothetical prote  
34 20 100.0 114 2 C81829 hypothetical prote  
35 20 100.0 114 2 C64511 hypothetical prote  
36 20 100.0 115 2 H72705 hypothetical prote  
37 20 100.0 116 2 H69023 conserved hypothet  
38 20 100.0 119 2 S42881 platelet basic pro  
39 20 100.0 120 2 D69209 hypothetical prote  
40 20 100.0 126 2 S35338 hypothetical prote  
41 20 100.0 127 2 C87640 hypothetical prote  
42 20 100.0 129 1 E69973 conserved hypothet  
43 20 100.0 129 2 F69502 conserved hypothet  
44 20 100.0 129 2 S75257 hypothetical prote  
45 20 100.0 130 2 B71025 hypothetical prote

## ALIGNMENTS

## RESULT 1

S24422

4-oxalocrotonate tautomerase (EC 5.3.2.-) dmpI [similarity] - Pseudomonas putida  
C;Species: Pseudomonas putida  
C;Date: 19-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 17-Mar-2000  
C;Accession: S24422

R;Shingler, V.; Powlowski, J.; Marklund, U.

J. Bacteriol. 174, 711-724, 1992

A;Title: Nucleotide sequence and functional analysis of the complete phenol/3,4-dimet  
A;Reference number: S24417; MUID:92121108; PMID:1732207

A;Accession: S24422

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-63 <SHI>

A;Cross-references: EMBL:X60835; NID:g45681; PIDN:CAA43229.1; PID:g45686

C;Superfamily: 4-oxalocrotonate tautomerase

C;Keywords: aromatic hydrocarbon catabolism; intramolecular oxidoreductase; isomerase  
F;2-63/Product: 4-oxalocrotonate tautomerase #status predicted <MAY>

F;2/Active site: Pro #status predicted

Query Match 100.0%; Score 20; DB 2; Length 63;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4

|||||

Db 9 IEGR 12

## RESULT 2

C70202

hypothetical protein BB0820 - Lyme disease spirochete

C;Species: Borrelia burgdorferi (Lyme disease spirochete)

C;Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 08-Oct-1999

C;Accession: C70202

R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh

son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Yu

; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.

Nature 390, 580-586, 1997

A;Authors: Smith, H.O.; Venter, J.C.

A;Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.

A;Reference number: A70100; MUID:98065943; PMID:9403685

A;Accession: C70202

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-66 <KLE>

A;Cross-references: GB:AE001180; NID:g2688755; PIDN:AAAC67174.1; PID:g268

A;Experimental source: strain B31

Query Match

Best Local Similarity 100.0%; Score 20; DB 2; Length 66;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
|  
|  
|  
|  
Db 63 IEGR 66

## RESULT 3

AI3396  
hypothetical protein BME1159 [imported] - Brucella melitensis (strain 16M)  
C:Species: Brucella melitensis  
C>Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 01-Feb-2002  
C:Accession: AI3396  
R:DelVecchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mujeer, C.; Los, T.; Ivanova,  
M.; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letessier,  
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis  
A:Reference number: AD3252; PMID:11756688  
A:Accession: AI3396  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-67 <KUR>  
A:Cross-references: GB:AE008917; PIDN:AAL52340.1; PID:g17983135; GSPDB:GN00190  
A:Experimental source: strain 16M  
C:Genetics:  
A:Gene: BME1159  
A:Map position: I

Query Match 100.0%; Score 20; DB 2; Length 67;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
|  
|  
|  
|  
Db 37 IEGR 40

## RESULT 4

D82686  
hypothetical protein XF1400 [imported] - Xylella fastidiosa (strain 9a5c)  
C:Species: Xylella fastidiosa  
C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
C:Accession: D82686  
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing  
Nature 406, 151-157, 2000  
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A:Reference number: A82515; MUID:20365717; PMID:10910347  
A:Note: for a complete list of authors see reference number A59328 below  
A:Accession: D82686  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-71 <SIM>  
A:Cross-references: GB:AE003971; GB:AE003849; NID:g9106402; PIDN:AAF84209.1; GSPDB:GN001  
A:Experimental source: strain 9a5c  
R:Simpson, A.J.G.; Reinach, P.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre, H  
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig  
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E  
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
A:Reference number: A59328  
A:Contents: annotation  
C:Genetics:  
A:Gene: XF1400

Query Match 100.0%; Score 20; DB 2; Length 71;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
|  
|  
|  
|  
Db 24 IEGR 27

## RESULT 5

T10269  
hydroxymethylglutaryl-CoA reductase (NADPH2) (EC 1.1.1.34) - sea-island cotton (fragm  
N:Alternate names: 3-hydroxy-3-methylglutaryl coenzyme A reductase  
C:Species: Gossypium barbadense (sea-island cotton)  
C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 03-Jun-2002  
C:Accession: T10269  
R:Joost, O.; Blanchini, G.; Bell, A.A.; Benedict, C.R.; Magill, C.W.  
Mol. Plant Microbe Interact. 8, 880-885, 1995  
A:Title: Differential induction of 3-hydroxy-3-methylglutaryl CoA reductase in two co  
A:Reference number: Z17010; MUID:96112742; PMID:8664497  
A:Accession: T10269  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-73 <JOJO>  
A:Cross-references: EMBL:S82272; NID:g1839560; PIDN:AAB47161.1; PID:g1839561  
C:Genetics:  
A:Note: Intron positions not resolved (incomplete sequence)  
C:Function:  
A:Description: catalyzes the NADP-dependent synthesis of mevalonate from 3-hydroxy-3--  
A:Pathway: isoprenoid biosynthesis  
C:Superfamily: hydroxymethylglutaryl-CoA reductase (NADPH)  
C:Keywords: coenzyme A; isoprenoid biosynthesis; NADP; oxidoreductase

Query Match 100.0%; Score 20; DB 2; Length 73;  
Best Local Similarity 100.0%; Pred. No. 3e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
|  
|  
|  
|  
Db 32 IEGR 35

## RESULT 6

H69460  
conserved hypothetical protein AF1689 - Archaeoglobus fulgidus  
C:Species: Archaeoglobus fulgidus  
C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 29-Sep-1999  
C:Accession: H69460  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E  
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
Nature 390, 364-370, 1997  
A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes,  
Smith, H.O.; Woese, C.R.; Venter, J.C.  
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch  
A:Reference number: A69250; MUID:98049343; PMID:9389475  
A:Accession: H69460  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-73 <KLE>  
A:Cross-references: GB:AE000987; GB:AE000782; NID:g2689310; PIDN:AAB89570.1; PID:g264  
C:Superfamily: conserved hypothetical protein MJ1122

Query Match 100.0%; Score 20; DB 2; Length 73;  
Best Local Similarity 100.0%; Pred. No. 3e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
|  
|  
|  
|  
Db 32 IEGR 35

## RESULT 7

S53481  
FUN57 protein - yeast (Saccharomyces cerevisiae)  
N:Alternate names: protein YAR029w

C:Species: Saccharomyces cerevisiae  
 C:Date: 05-May-1995 #sequence\_revision 01-Sep-1995 #text\_change 19-Apr-2002  
 C:Accession: S53481  
 R:Bussey, H.; Keng, T.; Storms, R.K.; Vo, D.; Zhong, W.; Fortin, N.; Barton, A.B.; Kabac  
 submitted to the EMBL Data Library, February 1994  
 A:Description: Sequencing of chromosome I of Saccharomyces cerevisiae: analysis of the 5  
 A:Reference number: S53458  
 A:Accession: S53481  
 A:Molecule type: DNA  
 A:Residues: 1-74 <BUS>  
 A:Cross-references: EMBL:L28920; NID:g1616966; PID:g456134; GSPDB:GN00001; MIPS:YAR029W  
 C:Genetics:  
 A:Gene: FUN57; MIPS:YAR029W  
 A:Cross-references: SGD:S0000077  
 A:Map position: 1R

Query Match 100.0%; Score 20; DB 2; Length 74;  
 Best Local Similarity 100.0%; Pred. No. 3e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IEGR 4  
 Db 31 IEGR 34

RESULT 8  
 AC2441  
 hypothetical protein asl5083 [imported] - Nostoc sp. (strain PCC 7120)  
 C:Species: Nostoc sp.  
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
 C:Accession: AC2441  
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi  
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata, S  
 DNA Res. 8, 205-213, 2001  
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana  
 A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Accession: AC2441  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-75 <KUR>  
 A:Cross-references: GB:BA000019; PIDN:BA06782.1; PID:g17134221; GSPDB:GN00179  
 A:Experimental source: strain PCC 7120  
 C:Genetics:  
 A:Gene: asl5083

Query Match 100.0%; Score 20; DB 2; Length 75;  
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IEGR 4  
 Db 22 IEGR 25

RESULT 9  
 AD2183  
 hypothetical protein asr3019 [imported] - Nostoc sp. (strain PCC 7120)  
 C:Species: Nostoc sp.  
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
 C:Accession: AD2183  
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi  
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata, S  
 DNA Res. 8, 205-213, 2001  
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana  
 A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Accession: AD2183  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-75 <KUR>  
 A:Cross-references: GB:BA000019; PIDN:BA06782.1; PID:g17132113; GSPDB:GN00179  
 A:Experimental source: strain PCC 7120

C:Genetics:  
 A:Gene: asr3019

Query Match 100.0%; Score 20; DB 2; Length 75;  
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
 Db 25 IEGR 28

RESULT 10

B83961  
 acyl-carrier protein acpA [imported] - Bacillus halodurans (strain C-125)  
 C:Species: Bacillus halodurans

C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 03-Jun-2002

C:Accession: B83961

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.;

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans

A:Reference number: A83650; MUID:20512592; PMID:11058132

A:Accession: B83961

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-77 <STO>

A:Cross-references: GB:AP001515; GB:BA000004; NID:g10174886; PIDN:BA06209.1; GSPDB:G

A:Experimental source: strain C-125

C:Genetics:

A:Gene: acpA

C:Superfamily: acyl carrier protein; acyl carrier protein homology

C:Keywords: carrier protein

Query Match 100.0%; Score 20; DB 2; Length 77;  
 Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
 Db 73 IEGR 76

RESULT 11

T49620

hypothetical protein B5022.20 [imported] - Neurospora crassa

C:Species: Neurospora crassa

C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000

C:Accession: T49620

R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatu

submitted to the Protein Sequence Database, May 2000

A:Reference number: Z25022

A:Accession: T49620

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-77 <SCH>

A:Cross-references: EMBL:AL355932; GSPDB:GN00116; NCSP:B5022.20

A:Experimental source: BAC clone B5022; strain OR74A

C:Genetics:

A:Gene: NCSP:B5022.20

A:Map position: 6

A:Introns: 35/2

Query Match 100.0%; Score 20; DB 2; Length 77;  
 Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
 Db 14 IEGR 17

RESULT 12

AE1946

2.



Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Query			ID	Description
	Score	Match	Length		
C 1	1103.2	30.6	1962	22 ABA44229	Human breast cell
C 2	1103.2	30.6	1962	22 AAK02966	Human brain expres
C 3	1103.2	30.6	1962	22 AAI12977	Probe #2910 for ge
C 4	1103.2	30.6	1962	22 AAI34336	Probe #3022 used t
C 5	1103.2	30.6	1962	22 AAI02895	Probe #2886 used t
C 6	1103.2	30.6	1962	24 ABS02920	Human genome-deriv
C 7	1101.8	30.6	1125	22 ABA49383	Human breast cell
C 8	1101.8	30.6	1125	22 AAI15732	Human brain expres
C 9	1101.8	30.6	1125	22 AAI22215	Probe #12148 for q

## ALIGNMENTS

## RESULT 1

ABA44229/c  
ID ABA44229 standard; DNA: 1962 BP.

ABA44229;

DT 01-FEB-2002 (first entry)

Human breast cell single exon nucleic acid probe #2924.

Human; microarray; single exon probe; gene expression; breast; disease; cancer; ss.

Homo sapiens.

WO200157271-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US00662.

04-FEB-2000; 2000US-0180312.

30-JUN-2000; 2000US-0608408.

21-SEP-2000; 2000US-0234687.

04-OCT-2000; 2000GB-0024263.

MOLE- ) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR:

WPI; 2001-496933/54.

New spatially-addressable set of single exon nucleic acid probes, useful for measuring gene expression in sample derived from human breast, comprises number of single exon nucleic acid probes -

Claim 1; SEQ ID NO 2924; 327bp + sequence listing; English.

The invention relates to a spatially-addressable set of single exon nucleic acid probes for measuring gene expression in a sample derived from human breast and BT 474 cells. The method involves contacting the probes with a collection of detectably labelled nucleic acids derived from mRNA of human breast, and then measuring the label bound to each probe of the microarray. The probes are useful for verifying the expression of regions of genomic DNA predicted to encode proteins. They are useful for gene discovery, and for determining predisposition and/or prognosing breast disease. Gene expression analysis is useful for assessing the toxicity of chemical agents on cells. The microarray of this invention presents a far greater diversity of probes for measuring gene expression, with far less bias than expressed sequence tag microarrays. The method is suitable for rapid production of functional information from genomic sequence. The present sequence is a single exon nucleic acid probe of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences).

Sequence 1962 BP; 545 A; 383 C; 350 G; 684 T; 0 other;

Query Match 30.6%; Score 1103.2; DB 22; Length 1962;  
 Best Local Similarity 99.7%; Pred. No. 5.6e-283;  
 Matches 1105; Conservative 0; Mismatches 3; Indels 0; Gaps

1581 AGATCATTTTCATGATTTGTATTAAGTAGTCCTTCTCCCAACACGACTCACATATATAT 1640  
 |||||  
 1364 AGATCATTTTCATGATTTGTATTAAGTAGTCCTTCTCCCAACACGACTCACATATATAT 1305  
 |||||  
 1641 TCCAAAAGGTGAATGCCTGCCCTATTTCAGCTTTGCAGAGTAGTCCCAAAAGAGGAGTTGA 1700  
 |||||  
 1304 TCCAAAAGGTGAATGCCTGCCCTATTTCAGCTTTGCAGAGTAGTCCCAAAAGAGGAGTTGA 1245  
 |||||  
 1701 AGGTGCCTATAGTGACATCAATAANTTCGACATGCTTCTATTGCCCAAGTGGAAAC 1760  
 |||||  
 1244 AGGTGCCTATAGTGACATCAATAANTTCGACATGCTTCTATTGCCCAAGTGGAAAC 1185  
 |||||  
 1761 CATCCAGCTCATATATGACAGTGGAAATGAATGCCACCACACATCATTTTAATCTCAGTT 1820  
 |||||  
 1184 CATCCACCTCATATGACAGTGGAAATGAATGCCACCACACATCATTTTAATCTCAGTT 1125  
 |||||  
 1821 TCAAAATACAAACGATGAAGAGTTCAAANTCCAGATACAGTGGAGTGACACACAGGGA 1880  
 |||||  
 1124 TCAAAATACAAACGATGAAGAGTTCAAANTCCAGATACAGTGGAGTGACACACAGGGA 1065  
 |||||  
 1881 GGGACCAAACTGAATTTCTACGCCCCAGAGGGTTACGAAATTTAGTTAGTCTCCATAAC 1940  
 |||||  
 1064 GGGACCAAACTGAATTTCTACAGCCCAGAGGGTTACGAAATTTAGTTAGTCTCCATAAC 1005  
 |||||  
 1941 ACTTCTTCCAGAGCGGAAATCCTTTTGGAGATATTTCCCAAGAAAACGCTTCCCGAA 2000  
 |||||  
 1004 ACTTCTTCCAGAGCGGAAATCCTTTTGGAGATATTTCCCAAGAAAACGCTTCCCGAA 945  
 |||||  
 2001 GTTTAAGACATGATGTTTAATCAACAAAGGAGAGCCCGAGGAAGAGTGAAATTCGCCCT 2060  
 |||||  
 944 GTTTAAGACATGATGTTTAATCAACAAAGGAGAGCCCGAGGAAGAGTGAAATTCGCCCT 885  
 |||||  
 2061 GGTAAATATTTTCACCTCTCCAAAAGCGCCAGTTGAGTCTCAATACCTTTGGATTTGCA 2120  
 |||||  
 884 GGTAAATATTTTCACCTCTCCAAAAGCGCCAGTTGAGTCTCAATACCTTTGGATTTGCA 825  
 |||||  
 2121 ACTGGAACATGGAGACATCACCTTTGAAAGGATACAAATTTGTCCAAAGTCAGCCTTGCTGAG 2180  
 |||||  
 824 ACTGGACATCGACATCATCTTTTCAAAGGATACAAATTTGTCCAAAGTCAGCCTTGCTGAG 765  
 |||||



**IX**

1760

1760



1124	Db	TTCAAAANTACAAACGATGAAGAGTTCAAATTCAGATAACAGTGGAGTGGACACACAGGGA	1065
1881	Qy	GGGACAAAACCTGAATTTCTACGGCCCAAGAGGGTTACGAAAAATTTAGTTAGTCCCAATAAC	1940
1064	Db	GGGACAAAACCTGAATTTCTACGGCCCAAGAGGGTTACGAAAAATTTAGTTAGTCCCAATAAC	1005
1941	Qy	ACTTCTTCACAGGCGGAAATCCCTTTTGGAGATATTCCTCAAGAAAAAGCGTTCCCGAA	2000
1004	Db	ACTTCTTCACAGGCGGAAATCCCTTTTGGAGATATTCCTCAAGAAAAAGCGTTCCCGAA	945
2001	Qy	GTTTAAGAGACATGATGTTAACTCAACAAGGAGAGCCACAGAAAGAGGTGAAAAATCCCTT	2060
944	Db	GTTTAAGAGACATGATGTTAACTCAACAAGGAGAGCCACAGAAAGAGGTGAAAAATCCCTT	885
2061	Qy	GGTAAATATTTCACTCTTCCAAAGACGCCCAGTTGAGTCTCAATACCTTGGATTGCA	2120
884	Db	GGTAAATATTTCACTCTTCCAAAGACGCCCAGTTGAGTCTCAATACCTTGGATTGCA	825
2121	Qy	ACTGGAACATGGAGACATCACTTTGAAAGAGTACAATTTGTCCAAAGTCAGCCTTGCTGAG	2180
824	Db	ACTGGAACATGGAGACATCACTTTGAAAGAGTACAATTTGTCCAAAGTCAGCCTTGCTGAG	765
2181	Qy	ATCATTTCTGATGAACCTCACAGCATGCTAAATAAATAACAAGTATATAACAGATGA	2240
764	Db	ATCATTTCTGATGAACCTCACAGCATGCTAAATAAATAACAAGTATATAACAGATGA	705
2241	Qy	AACAAATGACAGTTGGTGGCTCCACAGAAAAACAGAGTTTCATAAAAAGCATCTTGCCAAA	2300
704	Db	AACAAATGACAGTTGGTGGCTCCACAGAAAAACAGAGTTTCATAAAAAGCATCTTGCCAAA	645
2301	Qy	CAGCTTTAGGAGTGTCTGAAAGATTCGAGAGTTGACTTTTCTCGAGTGAAGTGTAAAGT	2360
644	Db	CAGCTTTAGGAGTGTCTGAAAGATTCGAGAGTTGACTTTTCTCGAGTGAAGTGTAAAGT	585
2361	Qy	GAATGTCATGACAGGGTCAAGATCCACCCCTGGACTTGGAGACACACAGCAAGATTAG	2420
584	Db	GAATGTCATGACAGGGTCAAGATCCACCCCTGGACTTGGAGACACACAGCAAGATTAG	525
2421	Qy	AGTGGAAACTCACACCCAAAAACCATAGGGCGGAAATGTGCAAAAGAAAGCCCCCATC	2480
524	Db	AGTGGAAACTCACACCCAAAAACCATAGGGCGGAAATGTGCAAAAGAAAGCCCCCATC	465
2481	Qy	TCGTATTGTTCCTACTGAAAGCCAGATGACAAAAGAAAGAAATTCACAGGGAAGAAA	2540
464	Db	TCGTATTGTTCCTACTGAAAGCCAGATGACAAAAGAAAGAAATTCACAGGGAAGAAA	405
2541	Qy	AGAGAACAGTAGAATGGAGAAATGCTGAAATCACATAGCGGTTACTGAGGTGTTACT	2600
404	Db	AGAGAACAGTAGAATGGAGAAATGCTGAAATCACATAGCGGTTACTGAGGTGTTACT	345
2601	Qy	TGGAAAGAAAGCTGCAGCATTTACACATAGTTACTTTGGGCTTTTGGCATGGGAAAAA	2660
344	Db	TGGAAAGAAAGCTGCAGCATTTACACATAGTTACTTTGGGCTTTTGGCATGGGAAAAA	285
2661	Qy	AAAGTATTTCTAGATCTTCTCGACGAA	2688
284	Db	AAAGTATTTCTAGATCTTCTCGACGTA	257

RESULT 6  
ABS02920/c  
ID ABS02920 standard: DNA: 1962 BP.

XX  
AC  
ABS02920;

19-AUG-2002 (first entry)

XX	Human genome-derived single exon probe from lung SEQ ID No 2911.
DE	
XX	
KW	Human; ds: single exon probe; asthma; lung cancer; COPD; ILD;
KW	chronic obstructive pulmonary disease; interstitial lung disease
KW	familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW	tuberculous sclerosis; Gaucher's disease; Niemann-Pick disease;

KW KW KW KW KW OS XX PN XX PD PF XX PR PR PR PR PA PI XX DR XX PT PT XX PS XX CC

Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis; pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome; pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia; primary ciliary dyskinesia; pulmonary hypertension; biliary membrane disease.

*Homo sapiens*.

WO200186003-A2.

15-NOV-2007

20 JAN 2001

[illegible]

26-MAY-2000; 2000US-207456P.

03-AUG-2000; 2000US-0632366.

21-SEP-2000; 2000US-236359P.  
27-SEP-2000; 2000US-236359P.

04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR,

WPI; 2002-114183/15.

Spatially-addressable set of single exon nucleic acid probes, used to measure gene expression in human lung samples -

Claim 1: SEO ID No 2911; 634pp; English.

The invention relates to a spatially-addressable set of single exon nucleic acid probes for measuring gene expression in a sample derived one from human lung comprising single exon nucleic acid probes having one of 12614 nucleic acid sequences mentioned in the specification, or their complements or the 12387 open reading frames derived from the 12614 probes. Also included are a microarray comprising the novel set of probes; the novel set of probes which hybridise at high stringency to a nucleic acid expressed in the human lung; measuring gene expression in a sample derived from human lung, comprising (a) contacting the array with a collection of detectably labeled nucleic acids derived from human lung mRNA, and (b) measuring the label detectably bound to each probe of the array; identifying exons in a eukaryotic genome, comprising (a) algorithmically predicting at least one exon from genomic sequences of the eukaryote; and (b) detecting specific hybridisation of detectably labeled nucleic acids from eukaryote lung mRNA, to a single exon probe, having a fragment identical to the predicted exon, the probe is included in the above mentioned microarray; assigning exons to a single gene, comprising (a) identifying exons from genomic sequence by the method above and (b) measuring the expression of each of the exons in several tissues and/or cell types using hybridisation to a single exon microarrays having a probe with the exon, where a common pattern of expression of the exons in the tissues and/or cell types indicates that the exons should be assigned to a single gene; a peptide comprising one of 12011 sequences, mentioned in the specification, or encoded by the probes/open reading frames (ORF). The probes are used for gene expression analysis, and for identifying exons in a gene, particularly using human lung derived mRNA and for the study of lung diseases such as asthma, lung cancer, chronic obstructive pulmonary disease (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinolysis, Karagazeri syndrome, fibrocystic pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension and hyaline membrane disease. The present sequence is a single exon probe of the invention.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftn.wipo.int/pub/published/bct\\_sequences](http://ftn.wipo.int/pub/published/bct_sequences).



XX SQ Sequence 1962 BP; 545 A; 383 C; 350 G; 684 T; 0 other;  
Query Match 30.6%; Score 1103.2; DB 24; Length 1962;  
Best Local Similarity 99.7%; Pred. No. 5.6e-283;  
Matches 1105; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1581 AGATCAATTTTCATGAATTTGATAAAGTGATCTTCTCCCAACACAGCTCAGTATATATAT 1640  
DB 1364 AGATCAATTTTCATGAATTTGATAAAGTGATCTTCTCCCAACACAGCTCAGTATATATAT 1305  
QY 1641 TCCAAAAGGTGAATGCTGCTTATTTTCAGCTTTTGCAGAGTAGCCAAAGAGGAGTTGA 1700  
DB 1304 TCCAAAAGGTGAATGCTGCTTATTTTCAGCTTTTGCAGAGTAGCCAAAGAGGAGTTGA 1245  
QY 1701 AGTGCTTATAGTGACAATCCCAATATTCGACATGCTTCTATTTGCCAACAAGTGGAAGAAC 1760  
DB 1244 AGTGCTTATAGTGACAATCCCAATATTCGACATGCTTCTATTTGCCAACAAGTGGAAGAAC 1185  
QY 1761 CATCCACCTCATATGACAGTGGATGAATGCCACCAATACATATTTTAATCTCAGCTT 1820  
DB 1184 CATCCACCTCATATGACAGTGGATGAATGCCACCAATACATATTTTAATCTCAGCTT 1125  
QY 1821 TCAAAATACAAAGCATGAAGTTCAAAATGCAGATACAGTGGAGGTGGACACAAGGGA 1880  
DB 1124 TCAAAATACAAAGCATGAAGTTCAAAATGCAGATACAGTGGAGGTGGACACAAGGGA 1065  
QY 1881 GGGACCAAACTGAATTTCTACGCGCCAGAGGGTTACGAAATTTAGTTAGTCCCATTAAC 1940  
DB 1064 GGGACCAAACTGAATTTCTACGCGCCAGAGGGTTACGAAATTTAGTTAGTCCCATTAAC 1005  
QY 1941 ACTTCTTCAGAGGGGGAATCCTTTTGGAGATATTCCTCCAAAGAAAGAGCTTCCCGAA 2000  
DB 1004 ACTTCTTCAGAGGGGGAATCCTTTTGGAGATATTCCTCCAAAGAAAGAGCTTCCCGAA 945  
QY 2001 GTTAAAGACATGATTTACTCAACAAGGAGAGCCAGGAGGTGAAATTTCCCT 2060  
DB 944 GTTAAAGACATGATTTACTCAACAAGGAGAGCCAGGAGGTGAAATTTCCCT 885  
QY 2061 GGTAAATATTTCACTCTTCCAAAGAGCCAGTTCAGTCTCAATACCTTGGATTGCA 2120  
DB 884 GGTAAATATTTCACTCTTCCAAAGAGCCAGTTCAGTCTCAATACCTTGGATTGCA 825  
QY 2121 ACTGGAACATGGAGACATCACTTTGAAAGGATACAAATTTGCCAAGTCAGCTTGGTGG 2180  
DB 824 ACTGGAACATGGAGACATCACTTTGAAAGGATACAAATTTGCCAAGTCAGCTTGGTGG 765  
QY 2181 ATCAATTTCTGATGACTCAGACATGCTTAAATATAAATCAAGCTATATAACAGATGA 2240  
DB 764 ATCAATTTCTGATGACTCAGACATGCTTAAATATAAATCAAGCTATATAACAGATGA 705  
QY 2241 AACAAATCACAGTTTGGTGGTCCACAGGAAACACAGTTTCATAAAGCATCTTGCCAAA 2300  
DB 704 AACAAATCACAGTTTGGTGGTCCACAGGAAACACAGTTTCATAAAGCATCTTGCCAAA 645  
QY 2301 CAGCTTAGAGTGTCTGAAAGATTCAGAGGTTTGACTTTTCCGCAAGTGAAGT 2360  
DB 644 CAGCTTAGAGTGTCTGAAAGATTCAGAGGTTTGACTTTTCCGCAAGTGAAGT 585  
QY 2361 GAATGGTTCATGACAGGTCAGATTCACCCCTGGACTTGGAGACACAGCAAGATTAG 2420  
DB 584 GAATGGTTCATGACAGGTCAGATTCACCCCTGGACTTGGAGACACAGCAAGATTAG 525  
QY 2421 AGTGGAACTCACACCAAAACCATAGCGGAAATGTGACAAAGAAAGCCCCCATC 2480  
DB 524 AGTGGAACTCACACCAAAACCATAGCGGAAATGTGACAAAGAAAGCCCCCATC 465  
QY 2481 TCTGATTTCTCACTGGAAAGCCAGATGACAAAGAAAGAAATCAAGGGAAGAA 2540  
DB 464 TCTGATTTCTCACTGGAAAGCCAGATGACAAAGAAAGAAATCAAGGGAAGAA 405  
QY 2541 AGAAGACATAGTAATGGAGAAATGCTGAAATCACATAGGCGTTACTGAAGTCTTACT 2600  
|||||

Db 404 AGAACAACAGTAGAATGGAGGAAATGCTGAAAATACATAGCGTTACTGAGTCTTACT 345  
QY 2601 TGAAGAAAGCTCGACATTTACACAGATAGTTACTTGGGCTTTTGGCCATGGGAGAAAA 2660  
Db 344 TGAAGAAAGCTCGACATTTACACAGATAGTTACTTGGGCTTTTGGCCATGGGAGAAAA 285  
QY 2661 AAAGTATTTCTAGATCTTCTTCGACGAA 2688  
Db 284 AAAGTATTTCTAGATCTTCTTCGACGTA 257  
|||||

RESULT 7  
ABA49383/c  
ID ABA49383 standard; DNA; 1125 BP.  
XX ABA49383;  
AC ABA49383;  
XX 01-FEB-2002 (first entry)  
DT Human breast cell single exon nucleic acid probe #8078.  
DE Human; microarray; single exon probe; gene expression; breast;  
KW disease; cancer; ss.  
XX Homo sapiens.  
OS WO200157271-A2.  
PN 09-AUG-2001.  
XX 30-JAN-2001; 2001WO-US00662.  
XX 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
PI WPI; 2001-496933/54.  
DR New spatially-addressable set of single exon nucleic acid probes,  
PT useful for measuring gene expression in sample derived from human  
PT breast, comprises number of single exon nucleic acid probes  
XX Claim 4; SEQ ID NO 8078; 327pp + sequence listing; English.  
The invention relates to a spatially-addressable set of single exon  
nucleic acid probes for measuring gene expression in a sample derived  
from human breast and BT 474 cells. The method involves contacting  
the probes with a collection of detectably labelled nucleic acids  
derived from mRNA of human breast, and then measuring the label  
bound to each probe of the microarray. The probes are useful for  
verifying the expression of regions of genomic DNA predicted to  
encode proteins. They are useful for gene discovery, and for  
determining predisposition and/or prognosing breast disease. Gene  
expression analysis is useful for assessing the toxicity of chemical  
agents on cells. The microarray of this invention presents a far greater  
diversity of probes for measuring gene expression, with far less bias  
than expressed sequence tag microarrays. The method is suitable for  
rapid production of functional information from genomic sequence. The  
present sequence is a single exon nucleic acid probe of the invention.  
Note: The sequence data for this patent did not form part of the  
printed specification, but was obtained in electronic format directly  
from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 1125 BP; 269 A; 232 C; 223 G; 401 T; 0 other;

Query Match		30.6%;	Score 1101.8;	DB 22;	Length 1125;	
Best Local Similarity		99.8%;	Pred. No. 1e-282;			
Matches 1103;		Conservative	0;	Mismatches	2;	Indels 0; Gaps 0;
QY	1581	AGATCAATTTTCATGAATGCTATAAAGTGTATCTCTCCCAACACGACCTCACTATATATAT	1640			
Db	1105	AGATCAATTTTCATGAATGCTATAAAGTGTATCTCTCCCAACACGACCTCACTATATATAT	1046			
QY	1641	TCCAAAGGTGATGCTGCTCTATTTTCAGCTTTTCGCAAGTAGCCAAAGAGGAGTTGA	1700			
Db	1045	TCCAAAGGTGATGCTGCTCTATTTTCAGCTTTTCGCAAGTAGCCAAAGAGGAGTTGA	986			
QY	1701	AGGTGCTATAGTACAAATCCAAATATTCGACATGCTTCTATTTGGCAACAAGTGGAAAC	1760			
Db	985	AGGTGCTATAGTACAAATCCAAATATTCGACATGCTTCTATTTGGCAACAAGTGGAAAC	926			
QY	1761	CATCCACCTCAATATGACAGTGGATGAATGCCACCAATACATTTTAAATCTCACGTT	1820			
Db	925	CATCCACCTCAATATGACAGTGGATGAATGCCACCAATACATTTTAAATCTCACGTT	866			
QY	1821	TCAAAATACAAAGCATGAAGAGTTCAAAATGCAGATAACAGTGGAGTGGACACAAAGGA	1880			
Db	865	TCAAAATACAAAGCATGAAGAGTTCAAAATGCAGATAACAGTGGAGTGGACACAAAGGA	806			
QY	1881	GGGACCAAACTGAATTTCTACGCTCCAGAGGTTACGAAATTTAGTTAGTCCCATAC	1940			
Db	805	GGGACCAAACTGAATTTCTACGCTCCAGAGGTTACGAAATTTAGTTAGTCCCATAC	746			
QY	1941	ACTCTTCCAGAGCGGAAATCTTTTGGAGTATTTCCCAAGAAAAACGCTTCCCGAA	2000			
Db	745	ACTCTTCCAGAGCGGAAATCTTTTGGAGTATTTCCCAAGAAAAACGCTTCCCGAA	686			
QY	2001	GTTTAAAGAGCATGATTTAACTCAACAGAGAGGAGCCAGGAGAGTGAATTTCCCT	2060			
Db	685	GTTTAAAGAGCATGATTTAACTCAACAGAGAGGAGCCAGGAGAGTGAATTTCCCT	626			
QY	2061	GGTAAATATTTCACTCTCTCCAAAGAGCGCCAGTGTGAGTCTCAATACCTTGGATTGCA	2120			
Db	625	GGTAAATATTTCACTCTCTCCAAAGAGCGCCAGTGTGAGTCTCAATACCTTGGATTGCA	566			
QY	2121	ACTGGAACATGGAGACATCACTTTGAAAGGATACAATTTGTCAAGTCAGCCTTGTGAG	2180			
Db	565	ACTGGAACATGGAGACATCACTTTGAAAGGATACAATTTGTCAAGTCAGCCTTGTGAG	506			
QY	2181	ATCATTTCTGATGACATCAAGCATGCTAAATAAATAAATAAAGTCAAGCTATAATACAGATGA	2240			
Db	505	ATCATTTCTGATGACATCAAGCATGCTAAATAAATAAATAAAGTCAAGCTATAATACAGATGA	446			
QY	2241	AACAAATGACAGTTTGGTGGCTCCACAGGAAAAACAGGTTTCATAAAAAGCATCTTGCCTAA	2300			
Db	445	AACAAATGACAGTTTGGTGGCTCCACAGGAAAAACAGGTTTCATAAAAAGCATCTTGCCTAA	386			
QY	2301	CAGCTTAGAGTGTCTGAAAGATTGCAGAGTTGACCTTTTCTCAGTGTGTAAGT	2360			
Db	385	CAGCTTAGAGTGTCTGAAAGATTGCAGAGTTGACCTTTTCTCAGTGTGTAAGT	326			
QY	2361	GAATGCTATGACAGGTTGAGATCCACCTCTGAGTCTGGAGACCCACAGCAAGATTAG	2420			
Db	325	GAATGCTATGACAGGTTGAGATCCACCTCTGAGTCTGGAGACCCACAGCAAGATTAG	266			
QY	2421	AGTGGAACTCACACCCCAAAAACCATAGGCGGAAATGTGACAAAGAAAGCCCCCATC	2480			
Db	265	AGTGGAACTCACACCCCAAAAACCATAGGCGGAAATGTGACAAAGAAAGCCCCCATC	206			
QY	2481	TCGTATTGTTCCACTGGAAAGCCAGATGACAAAGAAAGAAATTCACAGGGAAGAAAA	2540			
Db	205	TCGTATTGTTCCACTGGAAAGCCAGATGACAAAGAAAGAAATTCACAGGGAAGAAAA	146			
QY	2541	AGAGAACAGTAGAATGGAGGAAAAATGCTGAAAAATCACATAGGCGTTACTGAAGTGTACT	2600			
Db	145	AGAGAACAGTAGAATGGAGGAAAAATGCTGAAAAATCACATAGGCGTTACTGAAGTGTACT	86			
QY	2601	TGGAAGAAAGCTGCAGCATTACACAGATAGTTACTTGGGCTTTTGGCCATGGGAGAAAA	2660			
Db	86	TGGAAGAAAGCTGCAGCATTACACAGATAGTTACTTGGGCTTTTGGCCATGGGAGAAAA				
Db	85	TGGAAGAAAGCTGCAGCATTACACAGATAGTTACTTGGGCTTTTGGCCATGGGAGAAAA	26			
QY	2661	AAAGTATTTCTAGATCTTCTCGAC	2685			
Db	25	AAAGTATTTCTAGATCTTCTCGAC	1			
RESULT 8						
ID	AAK15732/c					
AC	AAK15732					
XX	05-NOV-2001 (first entry)					
DE	Human brain expressed single exon probe SEQ ID NO: 15723.					
KW	Human; brain expressed exon; gene expression analysis; probe;					
KW	microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;					
KW	epilepsy; cancer; ss.					
OS	Homo sapiens.					
PN	WO200157275-A2.					
PD	09-AUG-2001.					
XX	30-JAN-2001; 2001WO-US00667.					
PF	04-FEB-2000; 2000US-0180312.					
PR	26-MAY-2000; 2000US-0207456.					
PR	03-JUN-2000; 2000US-0608408.					
PR	30-AUG-2000; 2000US-0632366.					
PR	21-SEP-2000; 2000US-0234687.					
PR	27-SEP-2000; 2000US-0236359.					
PR	04-OCT-2000; 2000GB-0024263.					
XX	(MOLE-) MOLECULAR DYNAMICS INC.					
XX	Penn SG, Hanzel DK, Chen W, Rank DR;					
XX	WPI; 2001-483446/52.					
DR	Single exon nucleic acid probes for analyzing gene expression in human brains -					
PT	Example 4; SEQ ID NO: 15723; 650pp + Sequence Listing; English.					
PS	The present invention provides a number of single exon nucleic acid probes which are derived from genomic sequences expressed in the human brain. They can be used to measure gene expression in brain cell samples, which may enable the diagnosis and improved treatment of nervous system diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, epilepsy and cancers. The present sequence is one of the probes of the invention.					
XX	Sequence 1125 BP; 269 A; 232 C; 223 G; 401 T; 0 Other;					
XX	Query Match	30.6%;	Score 1101.8;	DB 22;	Length 1125;	
XX	Best Local Similarity	99.8%;	Pred. No. 1e-282;			
XX	Matches 1103;	Conservative	0;	Mismatches	2;	Indels 0; Gaps 0;
QY	1581	AGATCAATTTTCATGAATGCTATAAAGTGTATCTCTCCCAACACGACCTCACTATATATAT	1640			
Db	1105	AGATCAATTTTCATGAATGCTATAAAGTGTATCTCTCCCAACACGACCTCACTATATATAT	1046			
QY	1641	TCCAAAGGTGATGCTGCTCTATTTTCAGCTTTTCGCAAGTAGCCAAAGAGGAGTTGA	1700			
Db	1045	TCCAAAGGTGATGCTGCTCTATTTTCAGCTTTTCGCAAGTAGCCAAAGAGGAGTTGA	986			
QY	1701	AGGTGCTATAGTACAAATCCAAATATTCGACATGCTTCTATTTGGCAACAAGTGGAAAC	1760			
Db	985	AGGTGCTATAGTACAAATCCAAATATTCGACATGCTTCTATTTGGCAACAAGTGGAAAC	926			
QY	1761	CATCCACCTCAATATGACAGTGGATGAATGCCACCAATACATTTTAAATCTCACGTT	1820			
Db	925	CATCCACCTCAATATGACAGTGGATGAATGCCACCAATACATTTTAAATCTCACGTT	866			
QY	1821	TCAAAATACAAAGCATGAAGAGTTCAAAATGCAGATAACAGTGGAGTGGACACAAAGGA	1880			
Db	865	TCAAAATACAAAGCATGAAGAGTTCAAAATGCAGATAACAGTGGAGTGGACACAAAGGA	806			
QY	1881	GGGACCAAACTGAATTTCTACGCTCCAGAGGTTACGAAATTTAGTTAGTCCCATAC	1940			
Db	805	GGGACCAAACTGAATTTCTACGCTCCAGAGGTTACGAAATTTAGTTAGTCCCATAC	746			
QY	1941	ACTCTTCCAGAGCGGAAATCTTTTGGAGTATTTCCCAAGAAAAACGCTTCCCGAA	2000			
Db	745	ACTCTTCCAGAGCGGAAATCTTTTGGAGTATTTCCCAAGAAAAACGCTTCCCGAA	686			
QY	2001	GTTTAAAGAGCATGATTTAACTCAACAGAGAGGAGCCAGGAGAGTGAATTTCCCT	2060			
Db	685	GTTTAAAGAGCATGATTTAACTCAACAGAGAGGAGCCAGGAGAGTGAATTTCCCT	626			
QY	2061	GGTAAATATTTCACTCTCTCCAAAGAGCGCCAGTGTGAGTCTCAATACCTTGGATTGCA	2120			
Db	625	GGTAAATATTTCACTCTCTCCAAAGAGCGCCAGTGTGAGTCTCAATACCTTGGATTGCA	566			
QY	2121	ACTGGAACATGGAGACATCACTTTGAAAGGATACAATTTGTCAAGTCAGCCTTGTGAG	2180			
Db	565	ACTGGAACATGGAGACATCACTTTGAAAGGATACAATTTGTCAAGTCAGCCTTGTGAG	506			
QY	2181	ATCATTTCTGATGACATCAAGCATGCTAAATAAATAAATAAAGTCAAGCTATAATACAGATGA	2240			
Db	505	ATCATTTCTGATGACATCAAGCATGCTAAATAAATAAATAAAGTCAAGCTATAATACAGATGA	446			
QY	2241	AACAAATGACAGTTTGGTGGCTCCACAGGAAAAACAGGTTTCATAAAAAGCATCTTGCCTAA	2300			
Db	445	AACAAATGACAGTTTGGTGGCTCCACAGGAAAAACAGGTTTCATAAAAAGCATCTTGCCTAA	386			
QY	2301	CAGCTTAGAGTGTCTGAAAGATTGCAGAGTTGACCTTTTCTCAGTGTGTAAGT	2360			
Db	385	CAGCTTAGAGTGTCTGAAAGATTGCAGAGTTGACCTTTTCTCAGTGTGTAAGT	326			
QY	2361	GAATGCTATGACAGGTTGAGATCCACCTCTGAGTCTGGAGACCCACAGCAAGATTAG	2420			
Db	325	GAATGCTATGACAGGTTGAGATCCACCTCTGAGTCTGGAGACCCACAGCAAGATTAG	266			
QY	2421	AGTGGAACTCACACCCCAAAAACCATAGGCGGAAATGTGACAAAGAAAGCCCCCATC	2480			
Db	265	AGTGGAACTCACACCCCAAAAACCATAGGCGGAAATGTGACAAAGAAAGCCCCCATC	206			
QY	2481	TCGTATTGTTCCACTGGAAAGCCAGATGACAAAGAAAGAAATTCACAGGGAAGAAAA	2540			
Db	205	TCGTATTGTTCCACTGGAAAGCCAGATGACAAAGAAAGAAATTCACAGGGAAGAAAA	146			
QY	2541	AGAGAACAGTAGAATGGAGGAAAAATGCTGAAAAATCACATAGGCGTTACTGAAGTGTACT	2600			
Db	145	AGAGAACAGTAGAATGGAGGAAAAATGCTGAAAAATCACATAGGCGTTACTGAAGTGTACT	86			
QY	2601	TGGAAGAAAGCTGCAGCATTACACAGATAGTTACTTGGGCTTTTGGCCATGGGAGAAAA	2660			
Db	86	TGGAAGAAAGCTGCAGCATTACACAGATAGTTACTTGGGCTTTTGGCCATGGGAGAAAA				



Db	985	AGGTGCCTATAGTGACAAATCCAAATAATTCGACATGCTTCTATTGGCAACAAAGTGGAAAC	926
QY	1761	CATCCACCTCATATATGCACAGTGGAAATGAATGCCACACAATACATTTTAATCTCACGTT	1820
Db	925	CATCCACCTCATATGCACAGTGGAAATGAATGCCACACAATACATTTTAATCTCACGTT	866
QY	1821	TCAAAATACAAACGATGAAGAGTTCAAAATGCAGATAACAGTGGAGGTGGACACAAGGGA	1880
Db	865	TCAAAATACAAACGATGAAGAGTTCAAAATGCAGATAACAGTGGAGGTGGACACAAGGGA	806
QY	1881	GGGACCAAACTGAATTCACGCCCCAGAAAGGTTACGAAAATTTAGTTAGTCCCATAAC	1940
Db	805	GGGACCAAACTGAATTCACGCCCCAGAAAGGTTACGAAAATTTAGTTAGTCCCATAAC	746
QY	1941	ACTTCTTCCAGAGCGGAAATCCTTTTCAGGATATTCCTCAAGAAAACGCTTCCCGAA	2000
Db	745	ACTTCTTCCAGAGCGGAAATCCTTTTCAGGATATTCCTCAAGAAAACGCTTCCCGAA	686
QY	2001	GTTTAAGAGACATGATGTTAACTCAACAAGGAGAGCCAGGAAGAGTGGAAATTTCCCT	2060
Db	685	GTTTAAGAGACATGATGTTAACTCAACAAGGAGAGCCAGGAAGAGTGGAAATTTCCCT	626
QY	2061	GGTAAATATTTCACTCCTTCCAAAAGACGCCCAGTTGAGTCTCAATACCTTGGATTTGCA	2120
Db	625	GGTAAATATTTCACTCCTTCCAAAAGACGCCCAGTTGAGTCTCAATACCTTGGATTTGCA	566
QY	2121	ACTGGAACATGGAGACATCACTTTGAAGAGATACAAATTTGCGAAGTCAGCCTTGCTGAG	2180
Db	565	ACTGGAACATGGAGACATCACTTTGAAGAGATACAAATTTGCGAAGTCAGCCTTGCTGAG	506
QY	2181	ATCATTTCTGATGAATCTACAGCATGCTAAAATAAAAAATCAAGCTATAATAACAGATGA	2240
Db	505	ATCATTTCTGATGAATCTACAGCATGCTAAAATAAAAAATCAAGCTATAATAACAGATGA	446
QY	2241	AACAAATGACAGTTTGGTGGCTCCACAGAGAAAACAGGTTCTATAAAGCATCTTGCCAAA	2300
Db	445	AACAAATGACAGTTTGGTGGCTCCACAGAGAAAACAGGTTCTATAAAGCATCTTGCCAAA	386
QY	2301	CAGCTTAGGAGTGTCTGAAAGATTGCAGAGGTTGACATTTTTCCTGCAGTGAAGTAAAGT	2360
Db	385	CAGCTTAGGAGTGTCTGAAAGATTGCAGAGGTTGACATTTTTCCTGCAGTGAAGTAAAGT	326
QY	2361	GAATGTCATGACCCAGGTCAGATCCACCCCTGGACTTGGAGACCACAGAGATTTAG	2420
Db	325	GAATGTCATGACCCAGGTCAGATCCACCCCTGGACTTGGAGACCACAGAGATTTAG	266
QY	2421	AGTGGAAATCTCACACCCAAAACCATAGCGGAAATGTCACAAAAGAAAAGCCCCCATC	2480
Db	265	AGTGGAAATCTCACACCCAAAACCATAGCGGAAATGTCACAAAAGAAAAGCCCCCATC	206
QY	2481	TCGTATTGTTCCACTGGAAAGCCAGATGACAAAAGAAAAGAAAATCACAGGGAAGAAA	2540
Db	205	TCGTATTGTTCCACTGGAAAGCCAGATGACAAAAGAAAAGAAAATCACAGGGAAGAAA	146
QY	2541	AGAGAACAGTGAATGGAGGAAATGCTGAAAATCACATAGGCGTTACTGAAAGTGTACT	2600
Db	145	AGAGAACAGTGAATGGAGGAAATGCTGAAAATCACATAGGCGTTACTGAAAGTGTACT	86
QY	2601	TGGAAGAAAGCTGCAGCATTAACAGATAGTTACTTGGGCTTTTCCCATGGGAGAAA	2660
Db	85	TGGAAGAAAGCTGCAGCATTAACAGATAGTTACTTGGGCTTTTCCCATGGGAGAAA	26
QY	2661	AAAGTATTTCTAGATCTTCGAC	2685
Db	25	AAAGTATTTCTAGATCTTCGAC	1

## RESULT 9

RESULT 9  
AAI22215/c

ID AAI22215 standard; DNA; 1125 BP.

XX

AC

XX

12-OCT-2001 (first entry)

Probe #12148 for gene expression analysis in human cervical cell sample.

Probe; human; microarray; gene expression; cervical epithelial cell; cervical cancer; ss.

**Homo sapiens.**

WO200157278-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US00670.

04-FEB-2000: 2000US-0180312

26-MAY-2000; 2000US-0207456.  
30-JUN-2000; 2000US-0608408

03-AUG-2000; 2000US-0632366.  
27-SEP-2000: 2000US-0224697

27-SEP-2000; 2000US-0236359.

(MORE -) MORE POWER AND CONTROL -

[illegible][illegible]

CC/TACOCOT FOR (FBI)

analyzing gene expression in human cervical

Claim 25; SEQ ID No

The present invention relates to human single exon nucleic acids

(SENP). The present sequence is one such p from human HeLa cells. The SENPs can be us

microarray, which can be used for measuring human gene expression in a sample derived from human cervical epithelial cells. By measuring gene expression, the probes are therefore useful in grading and/or staging of diseases of the cervix, notably cervical cancer.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences).

Sequence 1125 BP; 269 A; 232 C; 223 G; 401 T; 0 other;

Query Match	30.6%	Score 1101.8;	DB 22;	Length 1125;
Best Local Similarity	99.8%	Pred. No. 1e-282;		

QY 1581 AGATCATTTTCATGAATTGTATAAAGTGATCCTTCTCCCAACCAGACTCACTATATTAT 1640

Db 1105 AGATCATTTTCATGAATTGTATAAGTGATCCTTCTCCCCAACCCAGACTCACTATATTAT 1046

QY 1641 TCCAAAAGGTGAATGCCCTGCCCTTATTTTCAGCCTTTGCAGAAAGTAGCCAAAAGAGGAGTTCA 1700

Db 1045 TCCAAAAGGTGAATGCCTGCCTTATTTTCAGCTTTGCAGAACTAGCCCAAAACACCCACCTTC

QY 1701 AGGTGCCTATAGTGACAAATCCAATAATTGACATCGCTTCTATTTCCTCCACACACCTCCACACAC 1700

Db  
985 AGGTGCCTATAGTGACAAATCCCAATAATTCGACATCCTTCTATTCCCCAGCATCCC

[illegible]

Db  
925 CATCCACCTCATATTCACACTCCTCAATCATCTCCAGCGATTGCTGGTGTGTTTGTTT

[illegible]

db 865

[illegible]

Db 805 GGGACCAAACTGAATCTTACAGCCAGAGGGTTACGAAAATTTAGTTAGTCCCATAC 746  
Qy 1941 ACTTCTTCCAGAGCGGAAATCCTTTTGGAGATATTTCCCAAGAAAAACGCTTCCCGAA 2000  
Db 745 ACTTCTTCCAGAGCGGAAATCCTTTTGGAGATATTTCCCAAGAAAAACGCTTCCCGAA 686  
Qy 2001 GTTTAAGAGACATGATGTTAACTCAACAGGAGAGCCAGGAGAGGTGAATAATCCCT 2060  
Db 685 GTTTAAGAGACATGATGTTAACTCAACAGGAGAGCCAGGAGAGGTGAATAATCCCT 626  
Qy 2061 GGTAAATATTTCACTCTCCAAAAGAGCCAGGTGAGTCTCAATACCTTGGATTGCA 2120  
Db 625 GGTAAATATTTCACTCTCCAAAAGAGCCAGGTGAGTCTCAATACCTTGGATTGCA 566  
Qy 2121 ACTGGAACATGGAGACATCACTTTGAAAGGATACAAATTTGTCGAAGTCTGCTGAG 2180  
Db 565 ACTGGAACATGGAGACATCACTTTGAAAGGATACAAATTTGTCGAAGTCTGCTGAG 506  
Qy 2181 ATGATTTCTGATGAACCTCAGAGATGCTAAATAAAAAATCAAGCTATATAACAGATGA 2240  
Db 505 ATGATTTCTGATGAACCTCAGAGATGCTAAATAAAAAATCAAGCTATATAACAGATGA 446  
Qy 2241 AACAAATGACAGTTTGGTGGTCCACAGAGAAACAGGTTTCATAAAGCATCTTGGCAAA 2300  
Db 445 AACAAATGACAGTTTGGTGGTCCACAGAGAAACAGGTTTCATAAAGCATCTTGGCAAA 386  
Qy 2301 CAGCTTAGGAGTGTCTGAAAGATGACAGAGTTGACATTTTCCCTGCAAGTGAATGTAAGT 2360  
Db 385 CAGCTTAGGAGTGTCTGAAAGATGACAGAGTTGACATTTTCCCTGCAAGTGAATGTAAGT 326  
Qy 2361 GAATGGTTCATGACAGAGGTGAGATCCACCCCTGGACTTGGAGACCACAGCAAGATTAG 2420  
Db 325 GAATGGTTCATGACAGAGGTGAGATCCACCCCTGGACTTGGAGACCACAGCAAGATTAG 266  
Qy 2421 AGTGAAACTCACACCCCAAAACCATAGGCGGAAATGTGACAAAGAAAGCCCCATC 2480  
Db 265 AGTGAAACTCACACCCCAAAACCATAGGCGGAAATGTGACAAAGAAAGCCCCATC 206  
Qy 2481 TCTGATTTCTTCCACTGGAAAGCCAGATGACAAAAAGAAAGAAATTCACAGGAAAGAAA 2540  
Db 205 TCTGATTTCTTCCACTGGAAAGCCAGATGACAAAAAGAAAGAAATTCACAGGAAAGAAA 146  
Qy 2541 AGAGAACAGTGAATGGAGGAAATGCTGAAAAATCACATAGGCGGTTACTGAGTGTACT 2600  
Db 145 AGAGAACAGTGAATGGAGGAAATGCTGAAAAATCACATAGGCGGTTACTGAGTGTACT 86  
Qy 2601 TGGAGAAAGCTGACAGATTACAGATAGTTACTTGGGCTTTTGGCATGGGAGAAAAA 2660  
Db 85 TGGAGAAAGCTGACAGATTACAGATAGTTACTTGGGCTTTTGGCATGGGAGAAAAA 26  
Qy 2661 AAGTATTTCTAGATCTTCTCGAC 2685  
Db 25 AAGTATTTCTAGATCTTCTCGAC 1

RESULT 10  
ID AAI47512/c  
XX AAI47512 standard; DNA; 1125 BP.  
AC AAI47512;  
XX AAI47512;  
DT 17-OCT-2001 (first entry)  
XX  
DE Probe #16198 used to measure gene expression in human placenta sample.  
DE  
XX Probe; microarray; human; placenta; antenatal diagnosis;  
KW genetic disorder; ss.  
XX  
XX Homo sapiens.  
XX WO200157272-A2.  
XX  
XX 09-AUG-2001.  
PD

XX 30-JAN-2001; 2001WO-US00663.  
PF  
XX 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
PA Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-48897/53.  
PI Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human placenta -  
DR Claim 25; SEQ ID No 16198; 654pp; English.  
XX The present invention relates to single exon nucleic acid probes (SENP).  
CC The present sequence is one such probe. The probes are useful for  
CC producing a microarray for predicting, measuring and displaying gene  
CC expression in samples derived from human placenta. The probes are useful  
CC for antenatal diagnosis of human genetic disorders.  
XX  
SQ Sequence 1125 BP; 269 A; 232 C; 223 G; 401 T; 0 other;

Query Match 30.6%; Score 1101.8; DB 22; Length 1125;  
Best Local Similarity 99.8%; Pred. No. 1e-282;  
Matches 1103; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1581 AGATCATTTTTCATGAATGTTATATAAGTATCTCTTCCCAACACAGACTCACTATATAT 1640  
Db 1105 AGATCATTTTTCATGAATGTTATATAAGTATCTCTTCCCAACACAGACTCACTATATAT 1046  
Qy 1641 TCCAAAGGTGAATGCTGCTCTTATTTTCAGATGATGATGATGATGATGATGATGATGAT 1700  
Db 1045 TCCAAAGGTGAATGCTGCTCTTATTTTCAGATGATGATGATGATGATGATGATGATGAT 986  
Qy 1701 AGTGCTCTATAGTGACAATCCCAATATTCGACATGCTTCTTATTTGCCAACCAAGTGGAAAC 1760  
Db 985 AGTGCTCTATAGTGACAATCCCAATATTCGACATGCTTCTTATTTGCCAACCAAGTGGAAAC 926  
Qy 1761 CATCCACCTCATATATGCACAGTGGAAATGAATGCCACCAATACATTTTAAATCTCACGTT 1820  
Db 925 CATCCACCTCATATATGCACAGTGGAAATGAATGCCACCAATACATTTTAAATCTCACGTT 866  
Qy 1821 TCAAAATACAAACGATGAAGAGTTCAAAATGCAGATGAATGATGATGATGATGATGATGATGAT 1880  
Db 865 TCAAAATACAAACGATGAAGAGTTCAAAATGCAGATGAATGATGATGATGATGATGATGATGAT 806  
Qy 1881 GGGACCAAAACTGAATTTACGGCCAGAGGGTTACGAAAATTTAGTTAGTCCCATAC 1940  
Db 805 GGGACCAAAACTGAATTTACGGCCAGAGGGTTACGAAAATTTAGTTAGTCCCATAC 746  
Qy 1941 ACTTCTTCCAGAGCGGAAATCCTTTTGGAGATATTTCCCAAGAAAAACGCTTCCCGAA 2000  
Db 745 ACTTCTTCCAGAGCGGAAATCCTTTTGGAGATATTTCCCAAGAAAAACGCTTCCCGAA 586  
Qy 2001 GTTTAAGAGACATGATGTTAACTCAACAGGAGAGCCAGGAGAGGTGAATAATTCCTCT 2060  
Db 685 GTTTAAGAGACATGATGTTAACTCAACAGGAGAGCCAGGAGAGGTGAATAATTCCTCT 626  
Qy 2061 GGTAAATATTTCACTCTCCAAAAGAGCCAGGTGAGTCTCAATACCTTGGATTGCA 2120  
Db 625 GGTAAATATTTCACTCTCCAAAAGAGCCAGGTGAGTCTCAATACCTTGGATTGCA 566  
Qy 2121 ACTGGAACATGGAGACATCACTTTGAAAGGATACAAATTTGTCGAAGTCTGCTGAG 2180  
Db 565 ACTGGAACATGGAGACATCACTTTGAAAGGATACAAATTTGTCGAAGTCTGCTGAG 506

QY 2181 ATCATTTCTGATGAACATCAGCATGCTAAATAAATAAATAAATCAAGCTATTAATAACAGATGA 2240  
DB 505 ATCAITTTCTGATGAACATCAGCATGCTAAATAAATAAATAAATCAAGCTATTAATAACAGATGA 446  
QY 2241 AACAAATGACAGTTTGGTGGCTCCACAGGAAACAGGTTTCATAAAGCATCTTCCCAA 2300  
DB 445 AACAAATGACAGTTTGGTGGCTCCACAGGAAACAGGTTTCATAAAGCATCTTCCCAA 386  
QY 2301 CAGCTTAGGAGTGTCTGAAGATTGCGAGAGTTGACTTTCTCGAGTGAGTGAAGT 2360  
DB 385 CAGCTTAGGAGTGTCTGAAGATTGCGAGAGTTGACTTTCTCGAGTGAGTGAAGT 326  
QY 2361 GAATGGTCATGACACAGGTCAGAAATCCACCCCTGGACTTGGAGACACAGCAAGATTAG 2420  
DB 325 GAATGGTCATGACACAGGTCAGAAATCCACCCCTGGACTTGGAGACACAGCAAGATTAG 266  
QY 2421 AGTGGAAACTCACACCCCAAAACCATAGGCGGAAATGTGACAAAAGAAAGCCGCCATC 2480  
DB 265 AGTGGAAACTCACACCCCAAAACCATAGGCGGAAATGTGACAAAAGAAAGCCGCCATC 206  
QY 2481 TCTGATTGTTCCACTGGAAGCCAGATGACAAAAGAAAGAAATCACAGGGAAGAAA 2540  
DB 205 TCTGATTGTTCCACTGGAAGCCAGATGACAAAAGAAAGAAATCACAGGGAAGAAA 146  
QY 2541 AGAAGACAGTAGAATGGAGGAAATGCTGAAATCACATAGCGCTTACTGAAGTGTACT 2600  
DB 145 AGAAGACAGTAGAATGGAGGAAATGCTGAAATCACATAGCGCTTACTGAAGTGTACT 86  
QY 2601 TGAAGAAAGCTGCAGCATTACACAGATAGTTACTTGGGCTTTTGGCATGGGAGAAA 2660  
DB 85 TGAAGAAAGCTGCAGCATTACACAGATAGTTACTTGGGCTTTTGGCATGGGAGAAA 26  
QY 2661 AAAGTATTTCCTAGATCTCTCGAC 2685  
DB 25 AAAGTATTTCCTAGATCTCTCGAC 1

RESULT 11  
AA107914/c  
ID AA107914 standard; DNA; 1125 BP.  
XX  
AC AA107914;  
XX  
DT 09-OCT-2001 (first entry)  
XX  
DE Probe #7905 used to measure gene expression in human breast sample.  
XX  
KW probe; human; breast disease; breast cancer; development disorder; ss;  
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.  
XX  
OS Homo sapiens.  
XX  
PN WO200157270-A2.  
XX  
PD 09-AUG-2001.  
XX  
PE 29-JAN-2001; 2001WO-US000661.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-476286/51.  
XX

PT Novel single exon nucleic acid probe used to measuring gene expression  
PT in a human breast -  
XX  
XX Claim 25; SEQ ID No 7905; 322pp; English.  
XX  
CC The present invention relates to novel single exon nucleic acid probes.  
CC The present sequence is one such probe. The probes are useful for  
CC measuring human gene expression in a human breast sample, where the probe  
CC hybridises at high stringency to a nucleic acid expressed in the human  
CC breast. The probes are useful for predicting, diagnosing, grading,  
CC staging, monitoring and prognosing diseases of the human breast,  
CC particularly those diseases with polygenic aetiology. The diseases  
CC include: breast cancer, disorders of development, inflammatory diseases  
CC of the breast, fibrocystic changes, proliferative breast disease and  
CC non-carcinoma tumours.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX  
SQ Sequence 1125 BP; 269 A; 232 C; 223 G; 401 T; 0 other;

Query Match 30.6%; Score 1101.8; DB 22; Length 1125;  
Best Local Similarity 99.8%; Pred. No. 1e-282;  
Matches 1103; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1581 AGATCATTTTCATGAATTTGATATAAAGTATCTTCTCCAAACAGACTACTATATAT 1640  
DB 1105 AGATCATTTTCATGAATTTGATATAAAGTATCTTCTCCAAACAGACTACTATATAT 1046  
QY 1641 TCCAAAGGTGAATGCTGCTTATTTTCAGCTTTGCGAAGTAGCCAAAAGAGAGTTGA 1700  
DB 1045 TCCAAAGGTGAATGCTGCTTATTTTCAGCTTTGCGAAGTAGCCAAAAGAGAGTTGA 986  
QY 1701 AGGTGCTTATAGTGACAAATCCAATAATTCGACATGCTTCTATTCGCCAACAGTGGAAAAC 1760  
DB 985 AGGTGCTTATAGTGACAAATCCAATAATTCGACATGCTTCTATTCGCCAACAGTGGAAAAC 926  
QY 1761 CATCCACCTCATATGACAGTGAATGAATGCCACCACAAATACATTTTAACTCTCAGTT 1820  
DB 925 CATCCACCTCATATGACAGTGAATGAATGCCACCACAAATACATTTTAACTCTCAGTT 866  
QY 1821 TCATAATACAAACGATGAAGATTCATAANTGCAGATACAGTGGAGTGACACAGGGA 1880  
DB 865 TCATAATACAAACGATGAAGATTCATAANTGCAGATACAGTGGAGTGACACAGGGA 806  
QY 1881 GGGACCAAACTGAATTTCTAGGCCCCAGAGGGTTACGAAAATTTAGTTAGTCCCATAC 1940  
DB 805 GGGACCAAACTGAATTTCTAGGCCCCAGAGGGTTACGAAAATTTAGTTAGTCCCATAC 746  
QY 1941 ACTTCTTCCAGAGCGGAAATCCTTTTGGAGTATTTCCCAAGAAAAGCTTCCCGAA 2000  
DB 745 ACTTCTTCCAGAGCGGAAATCCTTTTGGAGTATTTCCCAAGAAAAGCTTCCCGAA 686  
QY 2001 GTTTAAGAGACATGATGTTAACTCAACAGAGAGGCCAGAGAGTGAATAATCCCT 2060  
DB 685 GTTTAAGAGACATGATGTTAACTCAACAGAGAGGCCAGAGAGTGAATAATCCCT 626  
QY 2061 GGTAAATATTTCAGTCTTCCAAAAGACGCCAGTTGAGTCTCAATACCTTGGATTGCA 2120  
DB 625 GGTAAATATTTCAGTCTTCCAAAAGACGCCAGTTGAGTCTCAATACCTTGGATTGCA 566  
QY 2121 ACTGGAACATGGAGACATCACTTTGAAAGGATACAAATTTGTCCAGTACGCTTGTCTAG 2180  
DB 565 ACTGGAACATGGAGACATCACTTTGAAAGGATACAAATTTGTCCAGTACGCTTGTCTAG 506  
QY 2181 ATCATTTCTGATGAACATCAGCATGCTAAAATAAATAAATAAATCAAGCTATTAATAACAGATGA 2240  
DB 505 ATCATTTCTGATGAACATCAGCATGCTAAAATAAATAAATAAATCAAGCTATTAATAACAGATGA 446  
QY 2241 AACAAATGACAGTTTGGTGGCTCCACAGGAAACAGGTTTCATAAAGCATCTTCCCAA 2300  
DB 445 AACAAATGACAGTTTGGTGGCTCCACAGGAAACAGGTTTCATAAAGCATCTTCCCAA 386

QY 2301 CAGCTTAGAGTGTCTGAAAGATTGCGAGGTTGACTTTCTCTGCACTGCTGTAAGT 2360  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 385 CAGCTTAGAGTGTCTGAAAGATTGCGAGGTTGACTTTCTCTGCACTGCTGTAAGT 326  
 QY 2361 GAATGTCATGACAGGGTTCAGAAATCCACCCCTGGACTTGGAGACCACAGCAAGATTAG 2420  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 325 GAATGTCATGACAGGGTTCAGAAATCCACCCCTGGACTTGGAGACCACAGCAAGATTAG 266  
 QY 2421 AGTGAAGTACACACCCCAAAACCATAGCGGGAATGTGACAAAAGAAAGCCCCCATC 2480  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 265 AGTGAAGTACACACCCCAAAACCATAGCGGGAATGTGACAAAAGAAAGCCCCCATC 206  
 QY 2481 TCTGATTGTTCCACTGGAAGCCAGATGACAAAAGAAAGAAATACAGAGGCAAGAAA 2540  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 205 TCTGATTGTTCCACTGGAAGCCAGATGACAAAAGAAAGAAATACAGAGGCAAGAAA 146  
 QY 2541 AGAGAAGTGTAGTGTGAGGAAATGCTGAAATACATAGGCGTTTACTGAAAGTGTACT 2600  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 145 AGAGAAGTGTAGTGTGAGGAAATGCTGAAATACATAGGCGTTTACTGAAAGTGTACT 86  
 QY 2601 TGAAGAAGTGTGAGGAAATGCTGAAATACATAGGCGTTTACTGAAAGTGTACT 2660  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 85 TGAAGAAGTGTGAGGAAATGCTGAAATACATAGGCGTTTACTGAAAGTGTACT 26  
 QY 2661 AAGTATTCTCTAGATCTTCTCGAC 2685  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 25 AAGTATTCTCTAGATCTTCTCGAC 1

RESULT 12

ABSI5475/C  
 ID ABSI5475 standard; DNA; 1125 BP.  
 XX  
 AC ABSI5475;  
 XX  
 DT 19-AUG-2002 (first entry)  
 XX  
 DE Human genome-derived single exon probe ORF from lung SEQ ID No 15466.  
 XX  
 KW Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;  
 KW chronic obstructive pulmonary disease; interstitial lung disease;  
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagenen syndrome;  
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
 KW primary ciliary dyskinesia; pulmonary hypertension;  
 KW hyaline membrane disease; open reading frame; ORF.

OS Homo sapiens.  
 XX  
 PN WO200186003-A2.  
 XX  
 PD 15-NOV-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US000665.  
 XX  
 PR 04-FEB-2000; 2000US-180312P.  
 PR 26-MAY-2000; 2000US-207456P.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-063236P.  
 PR 21-SEP-2000; 2000US-234687P.  
 PR 27-SEP-2000; 2000US-236359P.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2002-114183/15.  
 XX  
 PT Spatially-addressable set of single exon nucleic acid probes, used to  
 measure gene expression in human lung samples -

XX  
 PS  
 XX  
 CC The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human lung comprising single exon nucleic acid probes having one of  
 CC 12614 nucleic acid sequences mentioned in the specification, or their  
 CC complements or the 12387 open reading frames derived from the 12614  
 CC probes. Also included are a microarray comprising the novel set of  
 CC probes; the novel set of probes which hybridize at high stringency to a  
 CC nucleic acid expressed in the human lung; measuring gene expression in a  
 CC sample derived from human lung, comprising (a) contacting the array with  
 CC a collection of detectably labeled nucleic acids derived from human lung  
 CC mRNA, and (b) measuring the label detectably bound to each probe of  
 CC the array; identifying exons in a eukaryotic genome, comprising  
 CC (a) algorithmically predicting at least one exon from genomic sequences  
 CC of the eukaryote; and (b) detecting specific hybridisation of detectably  
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
 CC having a fragment identical to the predicted exon, the probe is included  
 CC in the above mentioned microarray; assigning exons to a single gene,  
 CC comprising (a) identifying exons from genomic sequence by the method  
 CC above and (b) measuring the expression of each of the exons in several  
 CC tissues and/or cell types using hybridisation to a single exon  
 CC microarrays having a probe with the exon, where a common pattern of  
 CC expression of the exons in the tissues and/or cell types indicates that  
 CC the exons should be assigned to a single gene; a peptide comprising one  
 CC of 12011 sequences, mentioned in the specification, or encoded by the  
 CC probes/open reading frames (ORF). The probes are used for gene  
 CC expression analysis, and for identifying exons in a gene, particularly  
 CC using human lung derived mRNA and for the study of lung diseases  
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease  
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary  
 CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,  
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary  
 CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,  
 CC pulmonary alveolar proteinosis, Karagenen syndrome, fibrocystic  
 CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension  
 CC and hyaline membrane disease. The present sequence is a single exon  
 CC probe open reading frame of the invention.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1125 BP; 269 A; 232 C; 223 G; 401 T; 0 other;  
 SQ  
 Query Match 30.6%; Score 1101.8; DB 24; Length 1125;  
 Best Local Similarity 99.8%; Pred. No. 1e-282;  
 Matches 1103; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1581 AGATCAATTTTCATGAATGTGTATTAAGTATCTCTCCCAACACAGCTCCTACTATTAT 1640  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 1105 AGATCAATTTTCATGAATGTGTATTAAGTATCTCTCCCAACACAGCTCCTACTATTAT 1046  
 QY 1641 TCCAAAAGGTGAATGCCTGCCTTATTTTCAGCTTTTCAGAAAGTAGCCAAAAGAGGAGTTGA 1700  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 1045 TCCAAAAGGTGAATGCCTGCCTTATTTTCAGCTTTTCAGAAAGTAGCCAAAAGAGGAGTTGA 986  
 QY 1701 AGGTGCCTATAGTACAAATCCAAATTAATTCGACATGCTTCTATTGCCACAAGTGGAAAC 1760  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 985 AGGTGCCTATAGTACAAATCCAAATTAATTCGACATGCTTCTATTGGCAACAAAGTGGAAAC 926  
 QY 1761 CATCCACCTCATATGTCAGAGTGAATGAATGCCACCAATACATTTTAAATCTCACGTT 1820  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 925 CATCCACCTCATATGTCAGAGTGAATGAATGCCACCAATACATTTTAAATCTCACGTT 866  
 QY 1821 TCAAAATACAAACGATCAAGAGTTCAAAATGCAGATACAGTGGAGTGGACACAGGGA 1880  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 865 TCAAAATACAAACGATCAAGAGTTCAAAATGCAGATACAGTGGAGTGGACACAGGGA 806  
 QY 1881 GGGACCAAAACTGAATTTCTACGCCGCCAGAGGGTTACGAAAATTTAGTTAGTCCCATAC 1940  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 805 GGGACCAAAACTGAATTTCTACGCCGCCAGAGGGTTACGAAAATTTAGTTAGTCCCATAC 746



DB 492 GCCACTGGATGACATAGAAAAACCT 520  
|||||  
RESULT 14  
AAAA3989  
ID AAA43989 standard; cDNA: 537 BP.  
XX  
AC AAA43989;  
XX  
XX 21-AUG-2000 (first entry)  
XX  
DE Mouse secreted expressed sequence tag SEQ ID NO:564.  
XX  
XX Human; mouse; chicken; rat; secreted expressed sequence tag: SEST;  
KW expressed sequence tag; EST; probe; chemotactic; proliferative;  
KW immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;  
KW thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;  
KW antiviral; antidiabetic; antiasthmatic; vulnary; antiparkinsonian;  
KW antitumor; osteopathic; neuroprotective; neotropic; antipsoriatic;  
KW cerebroprotective; anticonvulsant; antidepressant; gene therapy;  
KW vaccine; autoimmune disorder; multiple sclerosis; allergic condition;  
KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;  
KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;  
KW central nervous system disorder; Alzheimer's disease; stroke;  
KW parkinson's disease; Huntington's disease; coagulation disorder;  
KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;  
KW tumour; infection; depression; psoriasis; ss.  
XX  
OS Mus musculus.  
XX  
XX WO200021991-A1.  
XX  
XX 20-APR-2000.  
XX  
XX 15-OCT-1999; 99WO-US24206.  
XX  
XX 15-OCT-1998; 98US-0104436.  
XX  
XX (GENY ) GENETICS INST INC.  
XX  
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;  
PI Merberg D, Treacy M, Bowman MR;  
XX  
XX WPI: 2000-317938/27.  
XX  
PT Isolated polynucleotides, and encoded proteins, comprising secreted  
PT expressed sequence tags (SESTs), useful for treating various disorders  
PT such as autoimmune, infectious, and central nervous system disorders -  
XX  
XX Claim 1; Page 352; 803pp; English.  
XX  
XX AAA43426 to AAA45925 represent specifically claimed secreted expressed  
XX sequence tags (SESTs), isolated from human, mouse, chicken and rat  
XX tissue sources. The SESTs can have a range of activities depending on  
XX the tissues they were isolated from. The activities include:  
XX chemotactic; proliferative; immunomodulatory; haematopoietic;  
XX chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;  
XX cytostatic; antibacterial; antifungal; antiviral; antidiabetic;  
XX antiasthmatic; vulnary; antitumor; osteopathic; neuroprotective;  
XX neotropic; antiparkinsonian; antipsoriatic; cerebroprotective;  
XX anticonvulsant; and antidepressant. The SESTs can be used for gene  
XX therapy and in vaccines. The SESTs are useful as probes for the  
XX identification and isolation of full-length cDNAs and genomic DNA  
XX molecules which correspond to the SESTs. Proteins encoded by the SESTs  
XX are useful in assays for determining biological activity and raising  
XX antibodies. They may be useful for treatment of autoimmune disorders  
XX (multiple sclerosis, insulin dependent diabetes), allergic conditions  
XX (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,  
XX osteoporosis, osteoarthritis, central nervous system disorders  
XX (Alzheimer's, Parkinson's, Huntington's, disease, stroke), coagulation  
XX disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's  
XX disease), tumours, bacterial, fungal or viral infections, depression and

CC psoriasis. AAA45926 to AAA45931 represent linker variants which are given  
CC in the exemplification of the present invention.  
XX  
SQ Sequence 537 BP; 168 A; 133 C; 105 G; 129 T; 2 other;  
Query Match 11.1%; Score 398.4; DB 21; Length 537;  
Best Local Similarity 87.5%; Pred. No. 1.2e-95;  
Matches 446; Conservative 0; Mismatches 63; Indels 1; Gaps 1;  
QY 2897 AATTGACAGAGCTCATTTCCAAAGTGGCGATCTCAGGATATGCGAGTTGCTTCT 2956  
DB 21 AATTGACAGAG-TTCATTTCACAGGTGGCTCAGCTCAGGACATCGAGTTGCTTCT 79  
QY 2957 CTTATTATTTATTCATGATGAGTGCAGCCACTGAATATATATCAAGTCTTTGATG 3016  
DB 80 CCTACTTTTATTCATGATGAGTGCAGCTCAGCCCTCAATATTTCCCAAGTCTTTCA 139  
QY 3017 AAGTTGATACAGATCAATCTGCTGCTGATATCAGCAGCTCAATATATTCACCAACTCAG 3076  
DB 140 AAGTAGACACAGACCAATCTGCTGCTGATAGGAAATCCGACACTGGCCACGA 199  
QY 3077 GAATTGACAGAGCTGCGTTAAGTTTCAGAGTTTGCAGCTCTGGAACACACATGCTAATAA 3136  
DB 200 GAATTGACAGAGCTGCGTTAAGTTTCAGAGTTTGCAGCTCTGGAACACACATGCTAATAA 259  
QY 3137 ATTGCTCAAAATGCTTCTGCTGCTGATATCAGCAGCTCAATATATTCACCAACTCAG 3196  
DB 260 ATTGCTCAAAATGCTTCTGCTGCTGATATCAGCTCAATCAACACATCCCACTGCTCAG 319  
QY 3197 AATCTTACTATGATCCGACCTGCCCGGTCACTAAAGTCTAGTAACAACTGTAAC 3256  
DB 320 AAGCATACTACGACCCCACTGCTCCGCTCACTAAGAGTCTTGTCACTCACTGTAAGC 379  
QY 3257 CAGTAAGTCAAAATCCCAAGCATATAGGACACAAATATAGTCTTGAATCA 3316  
DB 380 CAGTAAGTCAAAATCCCAAGCATATAGGACACAAATATAGTCTTGAATCA 439  
QY 3317 TGGGAGAGAGAAATGCTTTTAAATGATTCGTACCAAGCTTCTCATGTTGGTGGCC 3376  
DB 440 TGGGAGAGAGAAATGCTTTTCAAGATGATAGCAACCAATGTTCTCATGTTGGTGGCC 499  
QY 3377 AGTTGGATGACATAGAAAAACCTAGGA 3406  
DB 500 AGTTGGATGACATAGAAAAACCTAGGA 529  
RESULT 15  
ABQ56917  
ID -ABQ56917 standard; cDNA: 572 BP.  
XX  
AC ABQ56917;  
XX  
XX 02-AUG-2002 (first entry)  
XX  
XX Human colon cancer related nucleotide sequence SEQ ID NO:612.  
XX  
XX Human; colon cancer; cancer; tissue profiling; forensic; mapping;  
XX genetic analysis; diagnostic; antisense therapy; gene; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200229086-A2.  
XX  
XX 11-APR-2002.  
XX  
XX 02-OCT-2001; 2001WO-US30732.  
XX  
XX 02-OCT-2000; 2000US-237271P.  
XX  
XX (FARB ) BAYER CORP.  
XX  
XX Burgess C, Astle JH, Carroll E, Catino TJ, Dwivedi P, Molino GA;  
PI Thiaglingam A, Lewis ME;









GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 12, 2003, 22:48:45 ; Search time 85 Seconds  
(without alignments)  
12988.663 Million cell updates/sec

Title: US-10-023-888-1

Perfect score: 3600

Sequence: 1 atggagacagacacactctct.....atcgagacaattgaagtag 3600

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

- 1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/1/ina/PTCUS\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query Match	Length	DB ID	Description
1	71.4	2.0	5064	4 US-08-936-107A-8
2	67	1.9	7218	1 US-08-232-463-14
3	63	1.8	417	1 US-08-318-970B-46
4	62.4	1.7	660	2 US-08-902-516-1
5	61.4	1.7	393	1 US-08-137-117D-24
6	61.4	1.7	393	1 US-08-436-717-24
7	61.4	1.7	396	2 US-08-379-057-15
8	61.4	1.7	438	2 US-08-653-402B-11
9	58.2	1.6	393	1 US-08-137-117D-32
10	58.2	1.6	393	1 US-08-436-717-32
11	58.2	1.6	393	2 US-08-621-751A-9
12	58.2	1.6	393	4 US-08-579-378A-13
13	58.2	1.6	396	2 US-08-579-378A-17
14	58.2	1.6	396	2 US-08-483-636-1
15	58.2	1.6	396	2 US-08-483-632-1
16	56.6	1.6	122	1 US-07-634-278-78
17	56.6	1.6	122	1 US-08-477-728-78
18	56.6	1.6	122	1 US-08-474-040-78
19	56.6	1.6	122	1 US-08-487-200-78
20	56.6	1.6	122	4 US-08-484-537-78
21	56.6	1.6	393	1 US-07-634-278-66
22	56.6	1.6	393	1 US-08-477-728-66
23	56.6	1.6	393	1 US-08-474-040-66
24	56.6	1.6	393	1 US-08-487-200-66
25	56.6	1.6	393	4 US-08-484-537-66
26	56.6	1.6	428	3 US-08-589-939-4
27	56.6	1.6	8068	4 US-09-301-593-35

28	55.6	1.5	434	5 PCT-US93-11611-8
29	53.4	1.5	394	3 US-08-836-561-24
30	53	1.5	690	4 US-08-348-548-5
31	53	1.5	690	5 PCT-US95-15716-5
32	46.4	1.3	59	3 US-08-843-409-13
33	45.4	1.3	393	2 US-08-621-751A-13
34	44.6	1.2	4949	3 US-09-138-024-22
35	44.6	1.2	4949	4 US-09-404-066-22
36	44.4	1.2	115	1 US-07-634-278-106
37	44.4	1.2	115	1 US-08-477-728-106
38	44.4	1.2	115	1 US-08-474-040-106
39	44.4	1.2	115	1 US-08-487-200-106
40	44.4	1.2	115	4 US-08-484-537-106
41	44.4	1.2	117	1 US-08-458-516-18
42	44.4	1.2	120	1 US-07-634-278-26
43	44.4	1.2	120	1 US-08-477-728-26
44	44.4	1.2	120	1 US-08-474-040-26
45	44.4	1.2	120	1 US-08-487-200-26

## ALIGNMENTS

## RESULT 1

US-08-936-107A-8

; Sequence 8, Application US/08936107A

; Patent No. 6403306

; GENERAL INFORMATION:

; APPLICANT: Stephens, David S.

; APPLICANT: Swatley, John S.

; TITLE OF INVENTION: Serogroup-Specific Nucleotide Sequences

; TITLE OF INVENTION: in the Molecular Typing of Bacterial Isolates and the

; NUMBER OF SEQUENCES: 51

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.

; STREET: 5370 Manhattan Circle, Suite 201

; CITY: Boulder

; STATE: Colorado

; COUNTRY: US

; ZIP: 80303

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/936,107A

; FILING DATE: 23-SEP-1997

; CLASSIFICATION: 536

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/827,622

; FILING DATE: 09-APR-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Caruthers, Jennie M.

; REGISTRATION NUMBER: 34,464

; REFERENCE/DOCKET NUMBER: 77-97

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (303) 499-8080

; TELEFAX: (303) 499-8089

; INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 5064 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: not relevant

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 479..1597

; FEATURE:

; NAME/KEY: CDS

LOCATION: 1598..3236  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 3309..4052  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 4054..4917  
US-08-936-107A-8

Query Match 2.0%; Score 71.4; DB 4; Length 5064;  
Best Local Similarity 55.4%; Pred. No. 6.3e-10;  
Matches 160; Conservative 0; Mismatches 126; Indels 3; Gaps 1;

QY 945 TRACGAAGACTGAGTACTATCGGATCTATCGAGAGGCAATGACACCATGGTTCGGA 1004  
DB 2417 TAGAGATGAATAAATTCGCATAGCTTGGGAATAGTGGATCTTCATTCGAAA 2476

QY 1005 TATTTTCATGTACCAACGGGAGATTCATCCCTGGCTGAACTTGACATCCCTCGAGT 1064  
DB 2477 AATTTTATGTCTCTAAATGTGCTCCCGAGCATGGTGTAGATTAATACCCCTAAAT 2536

QY 1065 GACAATAGTAACACACAGGATGTTTTCGAAATTTGAGCCACTTGCCCTACCTTTAGTTC 1124  
DB 2537 TCAATGGTATATACGAAGAATATATGCAAA---AGTGGCTTCTCTACTTTTAGCTC 2593

QY 1125 ACCTGCTATTAAGTACGCTTCATCGCATCGAAGGCTGTCCGAAAGTATTATTACCT 1184  
DB 2594 ACATGCTATTGAACAGAGCTTGACCATATACCAAGAAATAGTAATATTATTACAG 2653

QY 1185 AAATGATGATGTATGTTGGGAAGTGTCTGGCCAGATGATTTTAC 1233  
DB 2654 CAATGAGACTTCTATTAACCAATGATTAAGAAATTTCTTC 2702

RESULT 2  
US-08-232-463-14/c  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149

INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZgpt-Fls  
US-08-232-463-14

Query Match 1.9%; Score 67; DB 1; Length 7218;  
Best Local Similarity 4.1%; Pred. No. 1.3e-08;  
Matches 16; Conservative 17; Indels 0; Gaps 0;

QY 2381 AGAATCCACCCCTGGACTTGGAGACACAGCAAGATTAGAGTGGAAACTCACACCCAAA 2440  
DB 1446 AGAATTTGGTACRR 1387

QY 2441 AAACCATAGGGGAAATGTGACAAAGAAAGCCCCCATCTCTGATTGTTCCACTGAAA 2500  
DB 1386 RRR 1327

QY 2501 GCCAGATGACAAAGAAAGAAATACAGGGAAGAAAGAAAGACAGATGAGTGGAGG 2560  
DB 1326 RRR 1267

QY 2561 AAAATGCTGAAATACATAGGCTTACTGAAGTGTACTTGGAGAAAGCTGCAGCAT 2620  
DB 1266 RRR 1207

QY 2621 ACACAGATAGTACTTGGGCTTTTGGCATGGGAAAGAAAGATTTCTAGATCTTC 2680  
DB 1206 RRR 1147

QY 2681 TCGACGAAGAAGTCAATTGAAGACACAATTTGGCTTACTTCACTGATGAGCAAGATAGAG 2740  
DB 1146 RRR 1087

QY 2741 CCAGATACAGAGATACATTTGCGAGATTCCTC 2775  
DB 1086 RRR 1052

RESULT 3  
US-08-318-970B-46  
Sequence 46, Application US/08318970B  
Patent No. 5589573  
GENERAL INFORMATION:  
APPLICANT: Hideaki HAGIWARA, et al.  
TITLE OF INVENTION: AMINO ACID SEQUENCES OF ANTI-IDIOTYPIC  
TITLE OF INVENTION: ANTIBODIES AGAINST ANTI-CANCER HUMAN MONOCLONAL ANTIBODY  
TITLE OF INVENTION: AND DNA BASE SEQUENCES ENCODING THOSE SEQUENCES  
NUMBER OF SEQUENCES: 48  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Law Office of Sherman and Shalloway  
STREET: 413 N. Washington Street  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22314  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage  
COMPUTER: Dell System 210; Intel 80 285 Microprocessor  
OPERATING SYSTEM: MS DOS 3.3  
SOFTWARE: Word Perfect, Version 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/318,970B  
FILING DATE: October 6, 1994  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Richard A. Steinberg  
REGISTRATION NUMBER: 26,588  
REFERENCE/DOCKET NUMBER: S-2371

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 549-2282  
TELEFAX: (703) 836-0106  
INFORMATION FOR SEQ ID NO: 46:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 417  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: mRNA  
HYPOTHETICAL: no  
ANTI-SENSE: no  
ORIGINAL SOURCE: mouse  
ORGANISM: mouse  
FEATURE:  
NAME/KEY: Clone 20KB1  
US-08-318-970B-46

Query Match 1.8%; Score 63; DB 1; Length 417;  
Best Local Similarity 100.0%; Pred. No. 4.1e-08;  
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGAGACAGACACTCTGCTATGGGTACTGCTGCTGGGTTCAGGTTCCACTGGT 60  
|||||  
Db 28 ATGGAGACAGACACTCTGCTATGGGTACTGCTGCTGGGTTCAGGTTCCACTGGT 87  
|||||  
QY 61 GAC 63  
|||  
Db 88 GAC 90

RESULT 4  
US-08-902-516-1  
Sequence 1, Application US/08902516  
Patent No. 5891432  
GENERAL INFORMATION:  
APPLICANT: Soo Hoo, William  
TITLE OF INVENTION: MEMBRANE-BOUND CYTOKINE COMPOSITIONS  
TITLE OF INVENTION: COMPRISING GM-CSF AND METHODS OF MODULATING AN IMMUNE  
TITLE OF INVENTION: RESPONSE USING SAME  
NUMBER OF SEQUENCES: 50  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CAMPBELL & FLORES, LLP  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: United States  
ZIP: 92121  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/902,516  
FILING DATE: 29-JUL-1997  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-IM 2442  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619)535-9001  
TELEFAX: (619)535-8949  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 660 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..660

US-08-902-516-1

Query Match 1.7%; Score 62.4; DB 2; Length 660;  
Best Local Similarity 86.2%; Pred. No. 7.6e-08;  
Matches 69; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 ATGGAGACAGACACTCTGCTATGGGTACTGCTGCTGGGTTCAGGTTCCACTGGT 60  
|||||  
Db 1 ATGGAGACAGACACTCTGCTATGGGTACTGCTGCTGGGTTCAGGTTCCACTGGT 60  
|||||  
QY 61 GACGAGATCAGGTAGTCC 80  
|||||  
Db 61 GACTATCCATATGATGTCC 80

RESULT 5  
US-08-137-117D-24  
Sequence 24, Application US/08137117D  
Patent No. 5795965  
GENERAL INFORMATION:  
APPLICANT: TSUCHIYA, Masayuki  
APPLICANT: SATO, Koh  
APPLICANT: BENDIG, Mary  
APPLICANT: JONES, Steven  
APPLICANT: SALDANHA, Jose  
TITLE OF INVENTION: RESHAPED HUMAN ANTIBODY TO HUMAN  
TITLE OF INVENTION: INTERLEUKIN-6 RECEPTOR  
NUMBER OF SEQUENCES: 158  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/137,117D  
FILING DATE: 20-DEC-1993  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/JP92/00544  
FILING DATE: 24-APR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-32084  
FILING DATE: 19-FEB-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 3-95476  
FILING DATE: 25-APR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: WEGNER, Harold C.  
REGISTRATION NUMBER: 25,258  
REFERENCE/DOCKET NUMBER: 53466/126/AAOK  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 393 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..393  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 1..393







; Sequence 13, Application US/08579378A  
; Patent No. 6210671  
; GENERAL INFORMATION:  
; APPLICANT: Co. Man Sung  
; TITLE OF INVENTION: Humanized Antibodies Reactive with  
; TITLE OF INVENTION: L-Selectin  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew  
; STREET: One Market Plaza, Steuart Tower, Suite 2000  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94105  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/579,378A  
; FILING DATE: 27-DEC-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/160,074  
; FILING DATE: 30-NOV-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/983,946  
; FILING DATE: 01-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 95112895.8  
; FILING DATE: 17-AUG-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 95114696.8  
; FILING DATE: 19-SEP-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Liebescheutz, Joe O.  
; REGISTRATION NUMBER: 37,505  
; REFERENCE/DOCKET NUMBER: 11823-002220  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-326-2400  
; TELEFAX: 415-326-2422  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 393 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cdna  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..393  
; US-08-579-378A-13

Query Match 1.6%; Score 58.2; DB 4; Length 393;  
Best Local Similarity 95.2%; Pred. No. 9e-07;  
Matches 60; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGGAGACAGACACACTCTGCTATGGGTACTGCTGCTGGTTCACAGTTCACACTGGT 60  
|||||  
Db 1 ATGGAGACAGACACAACTCTGCTATGGGTACTGCTGCTGGTTCACAGTTCACACTGGT 60  
|||||

QY 61 GAC 63  
|||  
Db 61 GAC 63

RESULT 13  
US-08-579-378A-17  
; Sequence 17, Application US/08579378A  
; Patent No. 6210671  
; GENERAL INFORMATION:  
; APPLICANT: Co. Man Sung

; TITLE OF INVENTION: Humanized Antibodies Reactive with  
; TITLE OF INVENTION: L-Selectin  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew  
; STREET: One Market Plaza, Steuart Tower, Suite 2000  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94105  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/579,378A  
; FILING DATE: 27-DEC-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/160,074  
; FILING DATE: 30-NOV-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/983,946  
; FILING DATE: 01-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 95112895.8  
; FILING DATE: 17-AUG-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 95114696.8  
; FILING DATE: 19-SEP-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Liebescheutz, Joe O.  
; REGISTRATION NUMBER: 37,505  
; REFERENCE/DOCKET NUMBER: 11823-002220  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-326-2400  
; TELEFAX: 415-326-2422  
; INFORMATION FOR SEQ ID NO: 17:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 393 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cdna  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..393  
; US-08-579-378A-17

Query Match 1.6%; Score 58.2; DB 4; Length 393;  
Best Local Similarity 95.2%; Pred. No. 9e-07;  
Matches 60; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGGAGACAGACACACTCTGCTATGGGTACTGCTGCTGGTTCACAGTTCACACTGGT 60  
|||||  
Db 1 ATGGAGACAGACACAACTCTGCTATGGGTACTGCTGCTGGTTCACAGTTCACACTGGT 60  
|||||

QY 61 GAC 63  
|||  
Db 61 GAC 63

RESULT 14  
US-08-483-636-1  
; Sequence 1, Application US/08483636  
; Patent No. 5914110  
; GENERAL INFORMATION:  
; APPLICANT: Holmes, Stephen D.  
; APPLICANT: Gross, Mitchell S.  
; APPLICANT: Sylvester, Daniel R.  
; TITLE OF INVENTION: Recombinant IL4 Antibodies Useful in  
; TITLE OF INVENTION: Treatment of IL4 Mediated Disorders





GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 12, 2003, 22:48:10 ; Search time 3185 Seconds  
(without alignments)  
18305.731 Million cell updates/sec

Title: US-10-023-888-1

Perfect score: 3600.

Sequence: 1 atggagacagacacactctct.....atcgagacaaattgaagtag 3600

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum Match 100%

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	771.8	21.4	1061	13	BM450679
2	730.4	20.3	1016	13	BM544697
3	648.6	18.0	984	12	BM250072
4	599.8	16.7	613	12	BG163659
5	589.2	16.4	616	14	BM790504
6	559.6	15.5	623	14	BQ305720

#### ALIGNMENTS

RESULT 1  
BM450679  
LOCUS  
DEFINITION  
AGENCOURT\_6394543 NIH\_MGC\_67 Homo sapiens cDNA Clone IMAGE:5494379  
5', mRNA sequence.  
ACCESSION  
BM450679  
VERSION  
BM450679.1  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 1061)  
NIH-MGC <http://mgc.ncbi.nlm.nih.gov/>.  
AUTHORS  
National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE  
Unpublished (1999)  
JOURNAL  
Contact: Robert Strausberg, Ph.D.  
COMMENT  
Email: [cgapbs-re@mail.nih.gov](mailto:cgapbs-re@mail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: L14W12119 row: j column: 12  
High quality sequence stop: 637.  
Location/Qualifiers  
1. .1061

7	559	15.5	567	12	BG291336
8	537.2	15.5	610	14	BQ307539
9	546.2	15.2	654	10	AW578427
10	546.2	15.2	654	10	AW604365
11	544.2	15.1	580	14	BQ306940
12	524.6	14.6	533	9	AI499228
13	514.4	14.3	572	12	BE736417
14	509	14.1	709	12	BF739893
15	507.2	14.1	673	12	BG085937
16	504.8	14.0	672	13	BI103956
17	480	13.3	929	14	BQ720582
18	480	13.3	934	14	BQ719691
19	478	13.3	981	14	BQ53143
20	468.8	13.0	600	9	AL135521
21	464.2	12.9	476	12	BF922316
22	455.4	12.7	693	9	AI664612
23	449.4	12.5	644	10	BE370268
24	440.6	12.2	484	14	BQ344941
25	434.8	12.1	1270	12	BG25753
26	431.2	12.0	596	13	BM364071
27	418.6	11.6	658	14	BQ305706
28	417.4	11.5	528	9	AA204698
29	414.4	11.5	482	12	BF931380
30	411.8	11.4	1037	14	BQ226717
31	397.4	11.0	637	13	BI660298
32	385.6	10.7	470	12	BF931580
33	378.8	10.5	459	12	BF931578
34	370.8	10.3	487	10	BB706869
35	369.6	10.3	1015	12	BF784322
36	350.8	9.7	356	14	N44886
37	343.8	9.6	815	13	BF757655
38	340.8	9.5	506	9	AI882177
39	335.8	9.3	442	12	BF566740
40	325.8	9.0	572	13	BM241944
41	320.2	8.9	550	17	AQ549054
42	309.2	8.6	322	12	BF934976
43	303	8.4	570	14	BM798105
44	280.8	7.8	579	12	BE695877
45	280.6	7.8	335	9	AA646238

BG291336 602387205  
BQ307539 MR0-BT400  
AW578427 RCL-CT024  
AW604365 RCL-CT024  
BQ306940 MR0-BT200  
AI499228 t08f11.x  
BE736417 601306591  
BF739893 704010.x  
BG085937 H3119H05-  
BI103956 602890235  
BQ720582 AGENCOURT  
BQ719691 AGENCOURT  
BQ53143 AGENCOURT  
AL135521 DRFP762K  
BF922316 QV2-NT014  
AI664612 uk26a06.y  
BE370268 601222515  
BQ344941 IL2-NT020  
BG25753 pgn1c.pk0  
BM364071 BS3000010  
BQ305706 MR0-BT200  
AA204698 zt87f01.r  
BF931380 IL2-NT020  
BQ226717 AGENCOURT  
BI660298 603302108  
BF931580 IL2-NT020  
BF931578 IL2-NT020  
BB706869 BB706869  
BF784322 602108193  
N44886 YZ21B04.f1  
BF757655 603027985  
AI882177 ul13d07.y  
BF566740 UI-R-BJ0p  
BM241944 K0624G04-  
AQ549054 KPCI-11-4  
BF934976 IL2-NT020  
BM798105 K-EST0081  
BE695877 RCL-CT024  
AA646238 vn14d02.r

QY 1278 AAACCTGTCGAGGGCTGCC 1298  
|||||  
Db 865 AAACCTGTCGAGGGCTGCC 885

RESULT 2  
BM544697 1016 bp mRNA linear EST 20-FEB-2002  
LOCUS NIH\_MGC\_124 Homo sapiens cDNA clone IMAGE:5727417  
DEFINITION 5' mRNA sequence.

ACCESSION BM544697  
VERSION BM544697.1 GI:18776157  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1016)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-r@mail.nih.gov  
Tissue Procurement: Invitrogen  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM12721 row: h column: 10  
High quality sequence stop: 690.

FEATURES  
Location/Qualifiers  
1..1016  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5727417"  
/clone\_lib="NIH\_MGC\_124"  
/tissue\_type="hippocampus"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pCMV-SPORT6; Site\_1: EcorV  
(destroyed); Site\_2: NotI; RNA source male hippocampus,  
age 27. Library is oligo-dT primed and directionally  
cloned (EcorV site is destroyed upon cloning). Average  
insert size 1.4 kb, insert size range 0.9-4 kb. Library is  
normalized and enriched for full-length clones and was  
constructed by C. Gruber (Invitrogen). Research Genetics  
tracking code 012."

BASE COUNT 302 a 226 c 224 g 260 t 4 others  
ORIGIN

Query Match 20.3%; Score 730.4; DB 13; Length 1016;  
Best Local Similarity 99.1%; Pred. No. 1.3e-160;  
Matches 743; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

QY 103 AGCCGAGATCAATACCATGTTTCTGTCATTCCTATAGACAAATATGCTGGAAAGTCC 162  
Db 105 AGCCGAGATCAATACCATGTTTCTGTCATTCCTATAGACAAATATGCTGGAAAGTCC 164

QY 163 TTTGAGATCGGCTTTGTCGTCATGCGGATTCGACGTTGTTTACACCTGGGTGAATGGC 222  
Db 165 TTTGAGATCGGCTTTGTCGTCATGCGGATTCGACGTTGTTTACACCTGGGTGAATGGC 224

QY 223 ACAGATCTTGAACACTACTGAAGAACTACAGCAGGTTCAGAGAACAGATGGAGGAGCAG 282  
Db 225 ACAGATCTTGAACACTACTGAAGAACTACAGCAGGTTCAGAGAACAGATGGAGGAGCAG 284

QY 283 AAAGCAATGAGAGAAATCCTTGGGAAACACACGGAACCTACTAAGAGAGTGAAG 342  
Db 285 AAAGCAATGAGAGAAATCCTTGGGAAACACACGGAACCTACTAAGAGAGTGAAG 344

QY 343 CAGTTAGAGTGTGCTAACACACTGCAATTAAGGTGCAATGCTGCTCGACCCAGCC 402  
|||||

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5494379"  
/clone\_lib="NIH\_MGC\_67"  
/tissue\_type="retinoblastoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: eye; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.75 kb. Library constructed by Life  
Technologies."

BASE COUNT 308 a 240 c 234 g 278 t 1 others  
ORIGIN

Query Match 21.4%; Score 771.8; DB 13; Length 1061;  
Best Local Similarity 95.9%; Pred. No. 2.6e-170; Indels 8; Gaps 3;  
Matches 826; Conservative 0; Mismatches 27;

QY 446 CTTCTTTTCATCTGCCAGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCA 505  
Db 25 CTTCTTTTCATCTGCCAGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCA 84

QY 506 ATGTCTCAGTGTGTTTGTGACGACTACTAAGGATGTTGAAGATGCCACCTCGACTGC 565  
Db 85 ATGTCTCAGTGTGTTTGTGACGACTACTAAGGATGTTGAAGATGCCACCTCGACTGC 144

QY 566 TTAAGGAATAGCAGACAGCAGATGAGGGGCTACTTGACACAGATAAAGAGTCC 625  
Db 145 TTAAGGAATAGCAGACAGCAGATGAGGGGCTACTTGACACAGATAAAGAGTCC 204

QY 626 CTGGATTAGTCTAATGCAAGATTTGGCTTTCTCTGAGTGGATTTCCACCAACATTCAGG 685  
Db 205 CTGGATTAGTCTAATGCAAGATTTGGCTTTCTCTGAGTGGATTTCCACCAACATTCAGG 264

QY 686 AACAATCACTAATAAACAATTTGCCAGAAATCTTCCCTAAAGTCAACCTGTTGC 745  
Db 265 AACAATCACTAATAAACAATTTGCCAGAAATCTTCCCTAAAGTCAACCTGTTGC 324

QY 746 AGTTGATTACAGAGCCAGTGTAGGCTTCTAAACTGAATTAACCCCAAGGATTTTCAAG 805  
Db 325 AGTTGATTACAGAGCCAGTGTAGGCTTCTAAACTGAATTAACCCCAAGGATTTTCAAG 384

QY 806 AATTGAATAGCAAACTAAGAGAACATGACCATGTAGTGAGAAAGACATGACCATAGTC 865  
Db 385 AATTGAATAGCAAACTAAGAGAACATGACCATGTAGTGAGAAAGACATGACCATAGTC 444

QY 866 CTGCATATTTATGATGATCTGAGGCGCATCAGCCAGTCTAAGCAGGATGAAGACATCT 925  
Db 445 CTGCATATTTATGATGATCTGAGGCGCATCAGCCAGTCTAAGCAGGATGAAGACATCT 504

QY 926 CTGCCAGTCTGTTTGAAGATAACGAAGAACTGAGGTACTCATTTGGGATCTATCGAGAGGC 985  
Db 505 CTGCCAGTCTGTTTGAAGATAACGAAGAACTGAGGTACTCATTTGGGATCTATCGAGAGGC 564

QY 986 ATGCACCATGGTTCGGAATATTTTCATTTGTCACCAACGGGAGATTCATCCTGGCTGA 1045  
Db 565 ATGCACCATGGTTCGGAATATTTTCATTTGTCACCAACGGGAGATTCATCCTGGCTGA 624

QY 1046 ACCTTGACATCTCGAGTGACAAATAGTAACACACAGGATGTTTTCGAAATTTGAGCC 1105  
Db 625 ACCTTGACATCTCGAGTGACAAATAGTAACACACAGGATGTTTTCGAAATTTGAGCC 684

QY 1106 ACTTGCTACCTTTAGTTCACCTGCTATTGAAAGTCACTGTCATCCGATCGAAGGCTGT 1165  
Db 685 ACTTGCTACCTTTAGTTCACCTGCTATTGAAAGTCACTGTCATCCGATCGAAGGCTGT 744

QY 1166 CCAGAGATTTATTTACCTAAATGATGATGCTATGTTGGG ---AAGGATGCTGSCCAG 1222  
Db 745 CCAGAGATTTATTTACCTAAATGATGATGCTATGTTGGGAAAGATGTTCTGGCCCCAG 804

QY 1223 ATGATTTTACAGTCAC -TCCAAAGGCCAGAAAGGTTTATTTGACATGGCC ---TGTGCC 1277  
Db 805 ATGATTTTACAGTCACCTCCAAAGGCCAGAAAGTTTATTTGACATGGCCCTGTGCCCA 864  
|||||

Db 345 CAGTTAGAGTGTGTTGCTTAACACACTGCATTAAGGTGCCAATGCTTCTCTGGACCCAGCC 404  
 QY 403 CTGCCGCCAACATCACCCTGAAGGACCTGCCATCTCTTTATTCCTCTCTTTTCAATCTGCC 462  
 Db 405 CTGCCGCCAACATCACCCTGAAGGACCTGCCATCTCTTTATTCCTCTCTTTTCAATCTGCC 464  
 QY 463 AGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGCTCTCAGTGTGTTGT 522  
 Db 465 AGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGCTCTCAGTGTGTTGT 524  
 QY 523 TTTGACAGTACTAAGGATGTTGAAGATGCCACTCTGGACTGCTTAAAGGAAATAGCAGA 582  
 Db 525 TTTGACAGTACTAAGGATGTTGAAGATGCCACTCTGGACTGCTTAAAGGAAATAGCAGA 584  
 QY 583 CAGACAGTATGGAGGGCTACTTGACAACAGATAAAGAGTCCCTGGATTAAGTGTGTTGT 642  
 Db 585 CAGACAGTATGGAGGGCTACTTGACAACAGATAAAGAGTCCCTGGATTAAGTGTGTTGT 644  
 QY 643 CAAGATTTGGCTTCTTCCAGTGGATTTCCACCAACATTCAGGAACCAAACTAAATAA 702  
 Db 645 CAAGATTTGGCTTCTTCCAGTGGATTTCCACCAACATTCAGGAACCAAACTAAATAA 704  
 QY 703 ACAAATTCGCCAGAAATCTTCTCTAAAGTCAAACTGTTGCAGTGTGTTATTCAGAGGCC 762  
 Db 705 ACAAATTCGCCAGAAATCTTCTCTAAAGTCAAACTGTTGCAGTGTGTTATTCAGAGGCC 764  
 QY 763 AGTGTAGCGCTTCTAAACCTGAATACCCCAAGGATTTCAAGATTTGAATTAAG-CAAA 821  
 Db 765 AGTGTAGCGCTTCTAAACCTGAATACCCCAAGGATTTCAAGATTTGAATTAAG-CAAA 824  
 QY 822 TAAGAGAACAATGACCATTTGATGGAAAGA 851  
 Db 825 TAAGAGAACAATGACCATTTGATGGAAAGA 854

RESULT 3  
 BG250072 984 bp mRNA linear EST 13-FEB-2001  
 LOCUS 602362161F1 NIH\_MGC\_90 Homo sapiens cDNA clone IMAGE:4470615 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BG250072  
 VERSION BG250072.1 GI:12759888  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 AUTHORS NIH-MGC http://mgs.nci.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLNL0288 row: 1 column: 16  
 High quality sequence stop: 670.  
 Location/Qualifiers  
 1. .984  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4470615"  
 /clone\_lib="NIH\_MGC\_90"  
 /tissue\_type="adenocarcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: liver; Vector: pCMV-SPORT6; Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally; oligo-dr primed.  
 Average insert size 1.7 kb. Library enriched for

FEATURES  
source

full-length clones and constructed by Life Technologies.  
 Note: this is a NIH\_MGC Library."  
 BASE COUNT 327 a 201 c 228 g 228 t  
 ORIGIN

Query Match 18.08; Score 648.6; DB 12; Length 984;  
 Best Local Similarity 92.8%; Pred. No. 2.1e-141;  
 Matches 747; Conservative 0; Mismatches 49; Indels 9; Gaps 6;

QY 2269 GAAAAACAGGTTCAATAAAAGCATCTTGCCAAACAGCTTTAGGAGTGTCTGAAAGATTGCAG 2328  
 Db 1 GAAAAACAGGTTCAATAAAAGCATCTTGCCAAACAGCTTTAGGAGTGTCTGAAAGATTGCAG 60  
 QY 2329 AGGTGTGACTTTTCTCGCAGTGAAGTGTAAAGTGAATGTCATGACCGGGTCAGAAATCCA 2388  
 Db 61 AGGTGTGACTTTTCTCGCAGTGAAGTGTAAAGTGAATGTCATGACCGGGTCAGAAATCCA 120  
 QY 2389 CCCCTGGACTTTGGAGACCCACAGCAAGATTTAGAGTGAAGTCAACCCCAAAACCATTA 2448  
 Db 121 CCCCTGGACTTTGGAGACCCACAGCAAGATTTAGAGTGAAGTCAACCCCAAAACCATTA 180  
 QY 2449 GCGCGAAATGTGACAAAGAAAGCCCACTCTCTGATTTGTTCCACTGGAAAGCCAGATG 2508  
 Db 181 GCGCGAAATGTGACAAAGAAAGCCCACTCTCTGATTTGTTCCACTGGAAAGCCAGATG 240  
 QY 2509 ACAAAGAAAGAAATCAACAGGGAAG-AAAAAGAGAACAGTAGAATGGAGGAAATGC 2567  
 Db 241 ACAAAGAAAGAAATCAACAGGGAAGAAAGAAAGAACAGTAGAATGGAGGAAATGC 300  
 QY 2568 TGAATAACACATAGCGCTTACTGAAAGTGTACTTGGAGAAAGCTGCAGCATTTACACAGA 2627  
 Db 301 TGAATAACACATAGCGCTTACTGAAAGTGTACTTGGAGAAAGCTGCAGCATTTACACAGA 360  
 QY 2628 TAGTTACTTGGGCTTTTGGCATGGGAGAAAGAAAGTATTTCTAGATCTTCTCGACGA 2687  
 Db 361 TAGTTACTTGGGCTTTTGGCATGGGAGAAAGAAAGTATTTCTAGATCTTCTCGACGA 420  
 QY 2688 AGAAGATCATTTGAAGACAAATTTGGCCTACTTCACTGATAGCAAGAAATAGAGCCAGATA 2747  
 Db 421 AGAAGATCATTTGAAGACAAATTTGGCCTACTTCACTGATAGCAAGAAATAGAGCCAGATA 480  
 QY 2748 CAAGAGATACATATTTGCAGATTTCCCTCAGATATGTAATAAATAATCTTAATAGCAAGTT 2807  
 Db 481 CAAGAGATACATATTTGCAGATTTCCCTCAGATATGTAATAAATAATCTTAATAGCAAGTT 540  
 QY 2808 TGGATTCACATCGCGGAAAGTCCCTGCTCAGATGCTTCACTGATGATGACCGGATTTGTTAT 2867  
 Db 541 TGGATTCACATCGCGGAAAGTCCCTGCTCAGATGCTTCACTGATGATGACCGGATTTGTTAT 600  
 QY 2868 GCAAGAACTGCAAGATATGTTCCCTGAAAGAAATTTGCAAGACGTCATTTCAAAAGTG-C 2926  
 Db 601 GCAAGAACTGCAAGATATGTTCCCTGAAAGAAATTTGCAAGACGTCATTTCAAAAGTGCC 660  
 QY 2927 GCCATCTGAGGATAT-GCAGTTTGGCTTCTCTTATTTT-TTATATCATCATGAGTGCAG 2983  
 Db 661 GCCATCTGAGGATATGCGAGTTTGGCTTCTCTTATTTTATGATGATGATGATGATGATG 720  
 QY 2984 TGAGCCACTGGAATATA---TCTCAAGTCTTTGATGAAGTTGATA-CAGATCAATCTGTT 3039  
 Db 721 TGAGCCACTGGAATATA---TCTCAAGTCTTTGATGAAGTTGATA-CAGATCAATCTGTT 780  
 QY 3040 GTCTTGTCTCAGAGAGAAATCCGAA 3064  
 Db 781 GTCTTGTCTGGAAGAAATTCGAA 805

RESULT 4  
 BG163659 613 bp mRNA linear EST 06-FEB-2001  
 LOCUS 60238839F1 NIH\_MGC\_89 Homo sapiens cDNA clone IMAGE:4446921 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BG163659  
 VERSION BG163659.1 GI:12670362

```

KEYWORDS      EST.
SOURCE         human.
ORGANISM       Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS        Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
TITLE          NIH-MGC http://mgi.nci.nih.gov/.
JOURNAL        National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT        Unpublished (1999)
               Contact: Robert Strausberg, Ph.D.
               Email: cgapbs-re@mail.nih.gov
               Tissue Procurement: ATCC
               cDNA Library Preparation: Life Technologies, Inc.
               cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
               DNA Sequencing by: Incyte Genomics, Inc.
               Clone distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               http://image.llnl.gov
               Plate: LLAM10226 row: n column: 10
               High quality sequence stop: 611.
               Location/Qualifiers
                 1..613
                 /organism="Homo sapiens"
                 /db_xref="taxon:9606"
                 /clone="IMAGE:4446921"
                 /clone_lib="NIH_MGC_89"
                 /tissue_type="hyponephroma, cell line"
                 /lab_host="DH10B (phage-resistant)"
                 /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                 Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                 Average insert size 1.3 kb. Library enriched for
                 full-length clones and constructed by Life Technologies.
                 Note: this is a NIH_MGC Library."
BASE COUNT    151 a 122 c 166 g 174 t
ORIGIN
Query Match      16.7%; Score 599.8; DB 12; Length 613;
Best Local Similarity 99.7%; Pred. No. 5.6e-130;
Matches 601; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1021 AACGGCAGATCCATCCTGGCTGACCTTGACATCTCCGAGTACATAGTACACAC 1080
      |||||||
Db 1 AACGGCAGATCCATCCTGGCTGACCTTGACATCTCCGAGTACATAGTACACAC 60
      |||||||
QY 1081 CAGGATGTTTTGCGAATTTGAGCCACTTGCCTACTCTTTAGTTTCACCTGCTATTGAAAGT 1140
      |||||||
Db 61 CAGGATGTTTTGCGAATTTGAGCCACTTGCCTACTCTTTAGTTTCACCTGCTATTGAAAGT 120
      |||||||
QY 1141 CAGGTTTCATCGATCGAAGGGCTGTCCCAAGAGTTTATTTACCTAAATGATGATGTCATG 1200
      |||||||
Db 121 CACATTCATCGATCGAAGGGCTGTCCCAAGAGTTTATTTACCTAAATGATGATGTCATG 180
      |||||||
QY 1201 TTTGGGAAGGATCTGCGCCAGATGATTTTACAGTCACCTCCAAAGGCCAGAGGTTTAT 1260
      |||||||
Db 181 TTTGGGAAGGATCTGCGCCAGATGATTTTACAGTCACCTCCAAAGGCCAGAGGTTTAT 240
      |||||||
QY 1261 TTGACATGGCTGTGCGCAAACTGTGCGGAGGCTGCCAGGTTCTCTGGATTAAAGGATGCG 1320
      |||||||
Db 241 TTGACATGGCTGTGCGCAAACTGTGCGGAGGCTGCCAGGTTCTCTGGATTAAAGGATGCG 300
      |||||||
QY 1321 TATTGTGACAGGCTTGTATTAATATCAGCTCGGATTTGGGATGGGGATGCTCTGGA 1380
      |||||||
Db 301 TATTGTGACAGGCTTGTATTAATATCAGCTCGGATTTGGGATGGGGATGCTCTGGA 360
      |||||||
QY 1381 AACATCGGAGGATGCTGTATATTGAGGAGTGGAGGCTGAGGATGAGTATGAGTGTGA 1440
      |||||||
Db 361 AACATCGGAGGATGCTGTATATTGAGGAGTGGAGGCTGAGGATGAGTATGAGTGTGA 420
      |||||||
QY 1441 CAGCCCTGGCAGTTTGGTGAGGAATAAAGAGTGTCTTACTGTATTAATCAGGAGTGTGGG 1500
      |||||||
Db 421 CAGCCCTGGCAGTTTGGTGAGGAATAAAGAGTGTCTTACTGTATTAATCAGGAGTGTGGG 480
      |||||||
QY 1501 AATTCCTGGCTCGCTGATAGTCTGTGTGACCAAGATGCAATGTCCTGCTGGGTTT 1560
      |||||||

Db 481 AATTCCTGGCTCGCTGATAGTCTGTGTGACCAAGATGCAATGTCCTGCTGGGTTT 540
      |||||||
QY 1561 GATGCTGGCAGTGTGGGCAAGATCATTTTCATGAATTTGTATAAGTATCTCTTCTCCCA 1620
      |||||||
Db 541 GATGCTGGCAGTGTGGGCAAGATCATTTTCATGAATTTGTATAAGTATCTCTTCTCCCA 600
      |||||||
QY 1621 AAC 1623
      |||
Db 601 AAC 603
      |||

RESULT 5
BM790504
LOCUS      BM790504
DEFINITION K-EST0070348 S22SNU16n1 Homo sapiens cDNA clone S22SNU16n1-7-F08
            5', mRNA sequence.
ACCESSION BM790504
VERSION    BM790504.1 GI:19138736
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (bases 1 to 616)
AUTHORS    Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
TITLE      21C Frontier Korean EST Project 2001
JOURNAL    Unpublished (2002)
COMMENT    Contact: Kim YS
            Korea Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
            Email: yongsung@mail.kribb.re.kr
            Plate: 7 row: F column: 08
            High quality sequence stop: 616.
            Location/Qualifiers
              1..616
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="S22SNU16n1-7-F08"
              /clone_lib="S22SNU16n1"
              /sex="F"
              /tissue_type="Ascites"
              /cell_type="lymphoblast-like"
              /lab_host="SNU-16"
              /note="Organ: Stomach; Vector: pT73-Pac; Site_1: EcoRI;
              Site_2: NotI; The S22SNU16 library was contributed by the
              Soares laboratory and it was constructed as described by the
              Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome
              Research 6(9): 791-806. RNA was prepared from harvested
              cells of SNU-16 culture. SNU-16 cell was obtained from
              Korean Cell Line Bank (KCLB). SNU-16 was established from
              ascitic fluids of Korean patients by Park J.G. et al.
              (1990), Cancer Res 50: 2773-2780."
BASE COUNT 155 a 117 c 168 g 176 t
ORIGIN
Query Match      16.4%; Score 589.2; DB 14; Length 616;
Best Local Similarity 99.2%; Pred. No. 1.7e-127;
Matches 613; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 1139 GTCAGTTTCATCGATCGAAGGGCTGCCAGAGTTTATTACCTAAATGATGATGTCGA 1198
      |||||||
Db 1 GTCACATTTCATCGATCGAAGGGCTGCCAGAGTTTATTACCTAAATGATGATGTCGA 60
      |||||||
QY 1199 TGTTTGGGAAGGATGCTGGCCAGATGATTTTTTACAGTCACCTCCAAAGGCCAAGAGTTT 1258
      |||||||
Db 61 TGTTTGGGAAGGATGCTGGCCAGATGATTTTTTACAGTCACCTCCAAAGGCCAAGAGTTT 120
      |||||||

```

QY 1259 ATTGACATGGCTGTGCGAACTGTGCCAGGCTGCCAGGTCTCTGGATTAAGGATG 1318  
 Db 121 ATTGACATGGCTGTGCCAACTGTGCCAGGCTGCCAGGTCTCTGGATTAAGGATG 180  
 QY 1319 GCTATTGTGACAGGCTTGTAAATATTCAGCCTGGGATGGGATGGTGGGATTCCTCTG 1378  
 Db 181 GCTATTGTGACAGGCTTGTAAATATTCAGCCTGGGATGGGATGGTGGGATTCCTCTG 240  
 QY 1379 GAACAGTGGAGGAGTCTCTATATTCAGGAGTGGAGTACTGGGATTAAGGATG 1438  
 Db 241 GAACAGTGGAGGAGTCTCTATATTCAGGAGTGGAGTACTGGGATTAAGGATG 300  
 QY 1439 GACAGCCCTGGGAGTGGTGGAGGAATAACAGTGTCTTACATGTAATCAGGATG 1498  
 Db 301 GACAGCCCTGGGAGTGGTGGAGGAATAACAGTGTCTTACATGTAATCAGGATG 360  
 QY 1499 CGAATTCCTGGCTGCTGATAGTCTGTGACCAAGCATGCAATCTCTGCTCTGGGT 1558  
 Db 361 CGAATTCCTGGCTGCTGATAGTCTGTGACCAAGCATGCAATCTCTGCTCTGGGT 420  
 QY 1559 TTGATGCTGGCGACTGTGGGCAAGATCATTTTCATGAATGTAATGAAGTGAATCTCTCC 1618  
 Db 421 TTGATGCTGGCGACTGTGGGCAAGATCATTTTCATGAATGTAATGAAGTGAATCTCTCC 480  
 QY 1619 CAACACAGACTCACTATATTTCCAAAGGTGAATGCTGCTTATTTACAGCTTTGCGAG 1678  
 Db 481 CAACACAGACTCACTATATTTCCAAAGGTGAATGCTGCTTATTTACAGCTTTGCGAG 540  
 QY 1679 AAGTAGCCAAAGAGAGTGAAGTGGCTATAGTACAAATCCAAATATTCAGATGCTT 1738  
 Db 541 AAGTAGCCAAAGAGAGTGAAGTGGCTATAGTACAAATCCAAATATTCAGATGCTT 598  
 QY 1739 CTATTGCCAACAGTGA 1756  
 Db 599 CTATTGCCAACAGTGA 616

RESULT 6  
 BQ305720/c  
 LOCUS  
 DEFINITION MRO-BT2002-290501-102-g09 BT2002 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BQ305720  
 VERSION BQ305720.1 GI:20835547  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 1 (bases 1 to 623)  
 Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.  
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 20202663  
 Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR0&t2=MR0-BT2002-290501-102-g09&t3=2001-05-29&t4=1)  
 290501-102-g09&t3=2001-05-29&t4=1)  
 Seq primer: puc 18 forward

FEATURES  
 source High quality sequence stop: 422.  
 Location/Qualifiers  
 1..623  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Br2002"  
 /dev\_stage="Adult"  
 /note="Organ: breast; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORSTES PCR (U.S. Letters Patent application No. 196 /716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."  
 BASE COUNT 160 a 128 c 137 g 198 t  
 ORIGIN  
 Query Match 15.5%; Score 559.6; DB 14; Length 623;  
 Best Local Similarity 97.5%; Pred. No. 1.5e-120;  
 Matches 579; Conservative 0; Mismatches 14; Indels 1; Gaps 1;  
 QY 2962 TTTTATTATCTCATGAGTGCAGTGCAGCCACTGAATATATCTCAAGTCTTTTGATGAAGTT 3021  
 Db 602 TGTATGATCTCATGAGT-CAGTGCAGCCACTGAATATATCTCAAGTCTTTTGATGAAGTT 544  
 QY 3022 GATACAGATCAATCTGTGTCTTGTCTGACAGAGAAATCCGACACTGGCTACAGAAAT 3081  
 Db 543 GATACAGATCAATCTGTGTCTTGTCTGACAGAGAAATCCGACACTGGCTACAGAAATG 484  
 QY 3082 CACGAACCTGCGTTAACTTTGACAGGATTTGACAGGCTGCGAACACATGCTAATAAATTCG 3141  
 Db 483 CACGGGCTGCGTTAACTTTGACAGGATTTGACAGGCTGCGAACACATGCTAATAAATTCG 424  
 QY 3142 TCRAAAATGCTTCTGCTGATATCAGCAGCTAAATAATATTCACCAACTCAGGAATCC 3201  
 Db 423 TCRAAAATGCTTCTGCTGATATCAGCAGCTAAAGAATATTCGCGCAACTCAGGAATCC 364  
 QY 3202 TACTATGATCCCACTGCCAGGCTCACTAAAAGTCTAGTAAACATGTAACACAGTA 3261  
 Db 363 TACTATGAGGCCCACTGCCAGGCTCACTAAAAGTCTAGTAAACATGTAACACAGTA 304  
 QY 3262 ACTGACAAAATCCACAAGCATATAAGGACAAAACAAATATAGTTTGAATCATGGGA 3321  
 Db 303 ACTGACAAAATCCACAAGCATATAAGGACAAAACCAATATAGTTTGAATCATGGGA 244  
 QY 3322 GAAGAGAAATCGGTTTAAATGATTCGTPACCAAGCTTTCTCATGTGTTGCCAGTTG 3381  
 Db 243 GAAGAGAAATCGGTTTAAATGATTCGTPACCAAGCTTTCTCATGTGTTGCCAGTTG 184  
 QY 3382 GATGACATAAGAAAACCCCTAGGAGTTTGTGCTGAATGACACATGACCAAT 3441  
 Db 183 GATGACATAAGAAAACCCCTAGGAGTTTGTGCTGAATGACACATGACCAAT 124  
 QY 3442 CATAAAGATGCTCAGACAGTGAAGGCTGTCTCAGGAGCTTCTATGAATCCATGTTCCCG 3501  
 Db 123 CATAAAGATGCTCAGACAGTGAAGGCTGTCTCAGGAGCTTCTATGAATCCATGTTCCCG 64  
 QY 3502 ATACCTTCCCAATTTGAATGCGCAAGAGATATCGAAACCGTTTCTCATATG 3555  
 Db 63 ATACCTTCCCAATTTGAATGCGCAAGAGATATCGAAACCGTTTCTCATATG 10

RESULT 7  
 BG291336  
 LOCUS  
 DEFINITION 602387205F1 NIH\_MGC\_93 Homo sapiens cDNA clone IMAGE:4516296 5', mRNA sequence.  
 ACCESSION BG291336  
 VERSION BG291336.1 GI:13049173  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 567)  
NIH-MGC <http://mhc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
Plate: LHAM10407 row: 1 column: 01  
High quality sequence stop: 567.

FEATURES  
source

1. 567  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4516296"  
/clone\_lib="NIH\_MGC\_93"  
/tissue\_type="transitional cell papilloma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: bladder; Vector: pCMV-SPORT6; Site\_1: Noti;  
Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 1.7 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: This is a NIH\_MGC Library."

BASE COUNT 192 a 123 c 110 g 142 t

ORIGIN

Query Match 15.5%; Score 559; DB 12; Length 567;  
Best Local Similarity 100.0%; Pred. No. 2.1e-120;  
Matches 559; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3040 GTCCTGTCGACAGAGAAATCCGACACTGGCTACCGAATTCACGAACTCCGCTTAAGT 3099  
DB 1 GTCCTGTCGACAGAGAAATCCGACACTGGCTACCGAATTCACGAACTCCGCTTAAGT 60

QY 3100 TTGCAGGATTTGACAGGCTCGAACACATGCTAATAATTCCTCAAAAATGCTTCCTGCT 3159  
DB 61 TTGCAGGATTTGACAGGCTCGAACACATGCTAATAATTCCTCAAAAATGCTTCCTGCT 120

QY 3160 GATATCAGCAGCTAAATAATTCACCAACTCAGGAATCCTACTATGATGCCAACCTG 3219  
DB 121 GATATCAGCAGCTAAATAATTCACCAACTCAGGAATCCTACTATGATGCCAACCTG 180

QY 3220 CCACGGTCACTAAAGTCTAGTAAACAACTGTAAACAGTAACTGACAAAATCCACAA 3279  
DB 181 CCACGGTCACTAAAGTCTAGTAAACAACTGTAAACAGTAACTGACAAAATCCACAA 240

QY 3280 GCATATAGGACAAACAAATATAGGTTGAAATCATCGGAGAGAGAAATCGCTTTT 3339  
DB 241 GCATATAGGACAAACAAATATAGGTTGAAATCATCGGAGAGAGAAATCGCTTTT 300

QY 3340 AAAATGATTCGACCAAGCTTCTCATGTTGGCCAGTTGGATGACATAAGAAAAAC 3399  
DB 301 AAAATGATTCGACCAAGCTTCTCATGTTGGCCAGTTGGATGACATAAGAAAAAC 360

QY 3400 CCTAGAGATTTGTTGGCTGAATGACAAATTTGACCAATCATAGATGCTCAGACA 3459  
DB 361 CCTAGAGATTTGTTGGCTGAATGACAAATTTGACCAATCATAGATGCTCAGACA 420

QY 3460 GTGAAGGCTGTTCTCAGGAGCTTCATGAATCCATGTTCCCATACCTTCCCAATTTGAA 3519  
DB 421 GTGAAGGCTGTTCTCAGGAGCTTCATGAATCCATGTTCCCATACCTTCCCAATTTGAA 480

QY 3520 CTGCCAAGAGATTCGAAACCGTTTCTTCATATGATGATGCTGAGGAGGCT 3579  
DB 481 CTGCCAAGAGATTCGAAACCGTTTCTTCATATGATGATGCTGAGGAGGCT 540

QY 3580 TATCGAGACAAATTTGAAT 3598

DB 541 TATCGAGACAAATTTGAAT 559

RESULT 8  
BQ307539/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

BQ307539 610 bp mRNA linear EST 16-MAY-2002  
MR0-BT4000-190601-202-f07 BT4000 Homo sapiens cDNA, mRNA sequence.  
BQ307539  
BQ307539.1 GI:20843892  
EST.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 610)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J., and  
Simpson,A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/gethtml2.pl?cl=MR0&t2=MR0-BT4000-190601-202-f07&t3=2001-06-19&t4=1>)  
Seq primer: puc 18 forward  
High quality sequence stop: 594.

FEATURES  
source

1. 610  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="BT4000"  
/dev\_stage="Adult"  
/note="Organ: breast; Vector: puc18; Site 1: SmaI; Site 2:  
SmaI; A mini-library was made by cloning products derived  
from ORESTES PCR (U.S. Letters Patent application No. 196  
/716 - Ludwig Institute for Cancer Research) profiles  
into the puc 18 vector. Reverse transcription of tissue  
mRNA and cDNA amplification were performed under low  
stringency conditions."

BASE COUNT 159 a 116 c 132 g 202 t 1 others

ORIGIN

Query Match 15.5%; Score 557.2; DB 14; Length 610;  
Best Local Similarity 99.0%; Pred. No. 5.6e-120;  
Matches 581; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

QY 2986 CAGCCACTGAATATATCTCAAGTC-TTTGATGAAGTTGATACAGATCAATCTGGTCTTT 3044  
DB 597 CAACCACTGAATATATCTCAAGTC-TTTGATGAAGTTGATACAGATCAATCTGGTCTTT 538

QY 3045 GTCTGACAGAGAAATCCGACACTGGCTACCGAATTCACGAACTCCGCTTAAGTTGCA 3104  
DB 537 GTCTGACAGAGAAATCCGACACTGGCTACCGAATTCACGAACTCCGCTTAAGTTGCA 478

QY 3105 GGATTTGACAGGCTTGGACACATGCTAATAAATTCCTCAAAAATGCTTCCTGCTGATAT 3164  
DB 477 GGATTTGACAGGCTTGGACACATGCTAATAAATTCCTCAAAAATGCTTCCTGCTGATAT 418

QY 3165 CAGCAGCTAAATAATATTTCCCAACTCAGGAATCCTACTATGATCCCCACCTGCCACC 3224



Db	417		CACGACGCTAAATAATATCCCAACTCAGGAATCTTACTATGATCCCCAACCTGNCACC	358
QY	3225	GGTCACATAAAGTGCTAGTAACAACACTGTAAACACGAGTAACGTGCACAAAATCCACAAAGCAT	3284	
Db	357	GGTCACATAAAGTGCTAGTAACAACACTGTAAACACGAGTAACGTGCACAAAATCCACAAAGCAT	298	
QY	3285	TAAGGACAAAAACAAATATATAGTTTGAATCATGGGAGAAGAAATATCGCTTTTAAAT	3344	
Db	297	TAAGGACAAAAACAAATATATAGTTTGAATCATGGGAGAAGAAATATCGCTTTTAAAT	238	
QY	3345	GATTCTGACCAACGTTTCTCATGTGGTTGCCAGTTGGATGACATATAGAAAAACCCCTAG	3404	
Db	237	GATTCTGACCAACGTTTCTCATGTGGTTGCCAGTTGGATGACATATAGAAAAACCCCTAG	178	
QY	3405	GAAGTTTGTGTTGGCTGAATGACAAACATTGACCAACATTAATAAGATGCTCAGACAGTGAA	3464	
Db	177	GAAGTTTGTGTTGGCTGAATGACAAACATTGACCAACATTAATAAGATGCTCAGACAGTGAA	118	
QY	3465	GGCTGTTCTCAGGACGTTCTATGAATCCATGTTCCCATACCTTCCCATAATTGAACATGCC	3524	
Db	117	GGCTGTTCTCAGGACGTTCTATGAATCCATGTTCCCATACCTTCCC-ATTGAACATGCC	59	
QY	3525	AAGAGATATCGAAACGTTTCTTCATATGCAATGACAGCTGCAGGAAT	3571	
Db	58	AAGAGATATCGAAACGTTTCTTCATATGCAATGACAGCTGCAGGAT	12	
RESULT	9			
AW578427				
LOCUS	AW578427	654 bp	mRNA	linear EST 16-MAR-2000
DEFINITION	RC1-CT0249-120100-022-d01 CT0249 Homo sapiens cDNA, mRNA sequence.			
ACCESSION	AW578427			
VERSION	AW578427.1	GI:72533476		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	HCGP (bases 1 to 654)			
TITLE	The FAPESP/LICR Human Cancer Genome Project			
JOURNAL	Unpublished (1999)			
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL ( <a href="http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC1&amp;t2=RC1-CT0249-120100-022-d01&amp;t3=2000-01-12&amp;t4=1">http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC1&amp;t2=RC1-CT0249-120100-022-d01&amp;t3=2000-01-12&amp;t4=1</a> ) Seq primer: puc 18 forward High quality sequence start: 6 High quality sequence stop: 517. Location/Qualifiers 1..654 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="CT0249" /dev_stage="Adult" /note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 /716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."			
FEATURES	source			
BASE COUNT	216 a	141 c	127 g	170 t
ORIGIN				





```

|||||
Db 248 CAGCAACTCCGTTAGTTTGAGGATTTGACAGGCTGGAACATGCTATAATATGC 307
QY 3142 TCAAAATGTTCTGCTGATATCAGCAGCTAAATAATATTCACCACTCAGGAATCC 3201
Db 308 TCAAAATGTTCTGCTGATATCAGCAGCTAAATAATATTCACCACTCAGGAATCC 367
QY 3202 TACTATGATCCCACTGCGCCAGGTCACATAAGTCTAGTAACAACTGTAACCCAGTA 3261
Db 368 TACTATGATCCCACTGCGCCAGGTCACATAAGTCTAGTAACAACTGTAACCCAGTA 427
QY 3262 ACTGACAAAATCACAAGCATATAAGGACAAAACAAATATAGTTTGAATCATGGGA 3321
Db 428 ACTGACAAAATCACAAGCATATAAGGACAAAACAAATATAGTTTGAATCATGGGA 487
QY 3322 GAAGAGAAATCGCTTTTAAATGATTCGTACCAAGTTTCATGTTGGTGGCCAGTTG 3381
Db 488 GAAGAGAAATCGCTTTTAAATGATTCGTACCAAGTTTCATGTTGGTGGCCAGTTG 547
QY 3382 GATGACATAGAAAACCTAGGAAGTTGTT 3414
Db 548 GATGACATAGAAAACCTAGGAAGTTGTT 580

RESULT 12
AI499228/c 533 bp mRNA linear EST 12-MAY-1999
LOCUS to08f11.x1 NCI_CGAP_Ut2 Homo sapiens cDNA clone IMAGE:2178477 3
DEFINITION similar to TR:Q61340 Q61340 KREISLER ;, mRNA sequence.
ACCESSION AI499228
VERSION AI499228.1 GI:4391210
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 533)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert length: 1973 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 412
POLYA-No.

FEATURES
source
Location/Qualifiers
1..533
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2178477"
/clone_lib="NCI_CGAP_Ut2"
/tissue_type="moderately-differentiated endometrial
adenocarcinoma, 3 pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site:1; SalI;
Site:2; NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.85 kb. Life Technologies catalog #:
11539-012"

BASE COUNT 116 a 113 c 114 g 188 t 2 others
ORIGIN
Query Match 14.6%; Score 524.6; DB 9; Length 533;
Best Local Similarity 98.9%; Pred. No. 2.5e-112;

```

```

Matches 527; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1978 CCRAAGAAAGAACGCTTCCCGAAGTTTAAAGAGACATGATGTTAACTCAAGAGAGAGCC 2037
Db 533 CCRAAGAAAGAACGCTTCCCGAAGTTTAAAGAGACATGATGTTAACTCAAGAGAGAGCN 474
QY 2038 CAGGAAGAGGTGAAAAATTCCTGTAATATTTCACTCTCTCCAAAGAGCGCCAGTTG 2097
Db 473 CAGGAAGAGGTGAAAAATTCCTGTAATATTTCACTCTCTCCAAAGAGCGCCAGTTG 414
QY 2098 AGTCTCAATACCTTTGGATTTGCAACTGGAACATGGAGACATCACTTTGAAAGGATCAAT 2157
Db 413 AGTCTCAATACCTTTGGATTTGCAACTGGAACATGGAGACATCACTTTGAAAGGATCAAT 354
QY 2158 TTGTCCAAGTCAGCCTTCTGAGATCATTTCTGATGAACACACAGCATGCTAAATATAA 2217
Db 353 TTGTCCAAGTCAGCCTTCTGAGATCATTTCTGATGAACACACAGCATGCTAAATATAA 294
QY 2218 AATCAAGCTATAATAACAGATGAAACAAATGACAGTTTGGTGGCTCCACAGGAAAAACAG 2277
Db 293 AATCAAGCTATAATAACAGATGAAACAAATGACAGTTTGGTGGCTCCACAGGAAAAACAG 234
QY 2278 GTTCATAAAAGCATCTTGCCAAACAGCTTAGAGAGTGTCTGAAAGATTGCAGAGGTTGACT 2337
Db 233 GTTCATAAAAGCATCTTGCCAAACAGCTTAGAGAGTGTCTGAAAGATTGCAGAGGTTGACT 174
QY 2338 TTTCTGTCAGTGAAGTAAAGTGAATGTCATGACACAGGTCAGAAATCCACCCCTGGAC 2397
Db 173 TTTCTGTCAGTGAAGTAAAGTGAATGTCATGACACAGGTCAGAAATCCACCCCTGGAC 114
QY 2398 TTGGAGACACAGCAAGATTTAGAGTGGAAACTCACACCCCAAAACCATAGGCGGAAT 2457
Db 113 TTGGAGACACAGCAAGATTTAGAGTGGAAACTCACACCCCAAAACCATAGGCGGAAT 54
QY 2458 GTGACAAAGAAAGCCCACTCTCTGATTTGTCACGTGGAAGAGCCAGATGAC 2510
Db 53 GTGACAAAGAAAGCCCACTCTCTGATTTGTCACGTGGAAGAGCCAGATGAC 1

RESULT 13
BE736417 572 bp mRNA linear EST 15-SEP-2000
LOCUS 601306591F1 NIH_MGC_39 Homo sapiens cDNA clone IMAGE:3640726 5',
DEFINITION mRNA sequence.
ACCESSION BE736417
VERSION BE736417.1 GI:10150409
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 572)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LNCM345 row: f column: 23
High quality sequence stop: 572.
Location/Qualifiers
1..572
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3640726"
/clone_lib="NIH_MGC_39"
/tissue_type="adenocarcinoma"

FEATURES
source

```

/lab\_host="DH10B (phage-resistant)"  
/notes="Organ: pancreas; Vector: pOTB7; Site 1: XhoI;  
Site 2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Library constructed  
by Ling Hong in the laboratory of Gerald M. Rubin  
(University of California, Berkeley) using ZAP-cDNA  
synthesis kit (Stratagene) and Superscript II RT (Life  
Technologies)."

BASE COUNT 197 a 124 c 114 g 136 t 1 others  
ORIGIN

Query Match 14.3%; Score 514.4; DB 12; Length 572;  
Best Local Similarity 98.2%; Pred. No. 6.1e-110;  
Matches 541; Conservative 0; Mismatches 7; Indels 3; Gaps 2;

QY 3048 TGACAGAGAAATCCGAACTGCTACCGAATTCAGCACTCCCGTTAAGTTTCAGGA 3107  
Db 1 TGACAGAGAAATCCGAACTGCTACCGAATTCAGCACTCCCGTTAAGTTTCAGGA 58  
QY 3108 TTTGACAGGCTCTGGAACACATGCTAATAAATGCTCAAAATGCTTCCTGCTGATATCAC 3167  
Db 59 ATTGACAGGCTCTGGAACACATGCTAATAA-TGCTCAAAATGCTTCCTGCTGATATCAC 117  
QY 3168 GCAGCTAAATAATATCCCACTCAGGAATCTCTACTATGATCCCACTGCCACCGGT 3227  
Db 118 GCAGCTAAATAATATCCCACTCAGGAATCTCTACTATGATCCCACTGCCACCGGT 177  
QY 3228 CACTAAAGTCTAGTAACCAACTGTAAACGAGTAAGTACCAAAATCCCAAGCATATAA 3287  
Db 178 CACTAAAGTCTAGTAACCAACTGTAAACGAGTAAGTACCAAAATCCCAAGCATATAA 237  
QY 3288 GGACAAAACAATATAGTTTGAATCATGGAGAGAAAGAAATCGCTTTTAAATGAT 3347  
Db 238 GGACAAAACAATATAGTTTGAATCATGGAGAGAAAGAAATCGCTTTTAAATGAT 297  
QY 3348 TCGTACCAAGTTCTCATGTGGTGGCCAGTTGGATGACATAGAAAGAAACCCCTAGGAA 3407  
Db 298 TCGTACCAAGTTCTCATGTGGTGGCCAGTTGGATGACATAGAAAGAAACCCCTAGGAA 357  
QY 3408 GTTTGTTGCTGGAATGACCAATGACCAATCATAAAGATGCTCAGACAGTGAAGGC 3467  
Db 358 GTTTGTTGCTGGAATGACCAATGACCAATCATAAAGATGCTCAGACAGTGAAGGC 417  
QY 3468 TGTCTCAGGACTCTATGATTCATGTTCCCATACCTTCCCAATTTGAATGCCAAG 3527  
Db 418 TGTCTCAGGACTCTATGATTCATGTTCCCATACCTTCCCAATTTGAATGCCAAG 477  
QY 3528 AGAGTATCGAAACCGTTTCTTCATATGATGATGAGTGGAGGCTTATCGAGA 3587  
Db 478 AGAGTATCGAAACCGTTTCTTCATATGATGATGAGTGGAGGCTTATCGAGA 537  
QY 3588 CAAATGGAAGT 3598  
Db 538 CAAATGGAAGT 548

RESULT 14  
BF739893/c  
LOCUS  
DEFINITION  
709 bp mRNA linear EST 10-JAN-2001  
7040d10.x1 NCI-CGAP\_Kid11 Homo sapiens cDNA clone IMAGE:3576619 3,  
similar to TR:Q9ULL2 Q9ULL2 KIAA1208 PROTEIN ; mRNA sequence.  
ACCESSION  
BF739893  
VERSION  
BF739893.1 GI:12066557  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 709)  
AUTHORS  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

JOURNAL  
COMMENT

Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaps-r@mail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL, send email to:  
info@image.llnl.gov

Seq primer: -40UP from Gibco  
High quality sequence stop: 460.

FEATURES  
source

1..709  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3576619"  
/clone\_lib="NCI\_CGAP\_Kid11"  
/lab\_host="DH10B"  
/notes="Organ: kidney; Vector: pOTB73D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Plasmid DNA from the normalized library NCI\_CGAP\_Kid3 was  
prepared, and ss circles were made in vitro. Following HAP  
purification, this DNA was used as tracer in a subtractive  
hybridization reaction. The driver was PCR-amplified cDNAs  
from a pool of 5,000 clones made from the same library  
(clones IDs 1322376-1323911, 1456007-1456775, and  
1500552-1502855). Subtraction by Bento Soares and M.  
Fatima Bonaldo."

BASE COUNT 174 a 141 c 135 g 259 t  
ORIGIN

Query Match 14.1%; Score 509; DB 12; Length 709;  
Best Local Similarity 98.1%; Pred. No. 1.1e-108;  
Matches 515; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2164 AAGTCAGCCTTGCTGAGATCATTTCTGATGAACCTCAGCAGCATGCTAAATAAATAATCAA 2223  
Db 709 AAGTCAGCCTTGCTGAGATCATTTCTGATGAACCTCAGCAGCATGCTAAATAAATAATCAA 650  
QY 2224 GCTATATAACAGATGAACAAATGACAGTTTGGTGGCTCCACAGGAAACAGGTTTCAT 2283  
Db 649 CTTATATAACAGATGAACAAATGACAGTTTGGTGGCTCCACAGGAAACAGGTTTCAT 590  
QY 2284 AAAAGCATCTTGCCAAACAGCTTAGGAGTCTCTGAAAGATTGACAGAGTTGACTTTTCCT 2343  
Db 589 AAAAGCATCTTGCCAAACAGCTTAGGAGTCTCTGAAAGATTGACAGAGTTGACTTTTCCT 530  
QY 2344 GCAGTGAAGTAAAGTGAATGCTATGACAGGGTCCAGAAATCCACCCCTGGACTTTGGAG 2403  
Db 529 GCAGTGAAGTAAAGTGAATGCTATGACAGGGTCCAGAAATCCACCCCTGGACTTTGGAG 470  
QY 2404 ACCACACAGATTTAGAGTGGAACTCAGCCCAACCAACCATAGCGGAAATGTGACA 2463  
Db 469 ACCACACAGATTTAGAGTGGAACTCAGCCCAACCAACCATAGCGGAAATGTGACA 410  
QY 2464 AAAAGAAAGCCCATCTCTGATTGTTTCCACTGGAAGCCAGATGACAAAGAAAGAAA 2523  
Db 409 AAAGAAAGCCCATCTCTGATTGTTTCCACTGGAAGCCAGATGACAAAGAAAGAAA 350  
QY 2524 ATCACAGGAAAGAAAGAAAGAAACAGTAGAATGGAGGAAATGCTGAAAATCACATAGGC 2583  
Db 349 ATCACAGGAAAGAAAGAAAGAAACAGTAGAATGGAGGAAATGCTGAAAATCACATAGGC 290  
QY 2584 GTTACTGAAGTGTACTTGGAGAAAGCTGCAGCATTCACAGATAGTTACTTGGGCTTT 2643  
Db 289 GTTACTGAAGTGTACTTGGAGAAAGCTGCAGCATTCACAGATAGTTACTTGGGCTTT 230  
QY 2644 TTGCCATGGGAGAAAGAAAGAAAGTATTTCCCTAGATCTTCTTCGACGAA 2688  
Db 229 TTGCCATGGGAGAAAGAAAGAAAGTATTTCCCAAGATCTTCTTCGACGTA 185

RESULT 15  
BG085937  
LOCUS  
DEFINITION H3119H05-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone  
H3119H05 5', mRNA sequence.  
ACCESSION BG085937  
VERSION 1  
KEYWORDS EST.  
SOURCE BG085937.1 GI:12568501  
ORGANISM house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 673)  
AUTHORS Kargul, G.J., Dudekula, D.B., Qian, Y., Lim, M.K., Jaradat, S.A., Tanaka  
, T.S., Carter, M.G. and Ko, M.S.H.  
TITLE Verification and initial annotation of NIA mouse 15K cDNA clone set  
JOURNAL Unpublished (2001)  
COMMENT Other\_ESTS: H3119H05-3  
Contact: George J. Kargul  
Laboratory of Genetics  
National Institute on Aging/National Institutes of Health  
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
Email: cdna@igsun.grc.nia.nih.gov  
This clone set has been freely distributed to the community. Please  
visit <http://igsun.grc.nia.nih.gov/cDNA/15k.html> for details.  
Plate: H3119 row: H column: 05  
Seq primer: -21M13 Reverse  
High quality sequence stop: 673  
POLYA=No.

## FEATURES

Location/Qualifiers  
1..673  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="niaEST:H3119H05-5"  
/db\_xref="taxon:10090"  
/clone="H3119H05"  
/clone\_lib="NIA Mouse 15K cDNA Clone Set"  
/sex="Clones arrayed from a variety of cDNA libraries"  
/dev\_stage="Clones arrayed from a variety of cDNA  
libraries"  
/lab\_host="DH10B"  
/note="Vector: pSPORT1; Site\_1: SalI; Site\_2: NotI; This  
clone is among a rearranged set of 15,247 clones from 11  
embryo cDNA libraries (including preimplantation stage  
embryos from unfertilized egg to blastocyst, embryonic  
part of E7.5 embryos, extraembryonic part of E7.5 embryos  
, and E12.5 female mesonephros/gonad) and one newborn  
ovary cDNA library. Average insert size 1.5 kb. All  
source libraries are cloned unidirectionally with Oligo(dt  
)-Not primers. References include: (1) Genome-wide  
expression profiling of mid-gestation placenta and embryo  
using a 15,000 mouse developmental cDNA microarray, 2000,  
Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2)  
Large-scale cDNA analysis reveals phased gene expression  
patterns during preimplantation mouse development, 2000,  
Development, 127: 1737-1749; (3) Genome-wide mapping of  
unselected transcripts from extraembryonic tissue of  
7.5-day mouse embryos reveals enrichment in the t-complex  
and under-representation on the X chromosome, 1998, Hum  
Mol Genet 7: 1967-1978."  
BASE COUNT 207 a 167 c 136 g 163 t  
ORIGIN

Query Match 14.1%; Score 507.2; DB 12; Length 673;  
Best Local Similarity 88.3%; Pred. No. 3e-108;  
Matches 551; Conservative 0; Mismatches 73; Indels 0; Gaps 0;  
QY 2975 TGAGTGCAGTCGACGACCTGATATCTCAAGTCTTTGATGACGATGATCAAT 3034  
|||||  
Db 1 TGAGTGCAGTCGACGACCTGATATCTCAAGTCTTTGATGACGATGATCAAT 3034  
|||||  
QY 3035 CTGGTGTCTTGTCTGACAGAGAAATCCGACACTGGCTTACCAGAAATTCACGAACTGCCGT 3094

Db 61 CTGGTGTCTTGTCTGATAGGGAATCCGACACTGGCCACGAGAAATTCACGACCTACCTT 120  
QY 3095 TAAGTTTCCAGGATTTGACAGGCTCTGGAACACATGCTTAATAAATGCTCAAAAATGCTTC 3154  
|||||  
Db 121 TAAGCTTGCAGGATTTGACAGGTTTGGAAACACATGTTTAATAAATGCTCAAAAATGCTTC 180  
QY 3155 CTGCTGATATCACGACGCTAAATATATTCACCACAACTCAGGAATCCTTACTATGATCCCA 3214  
|||||  
Db 181 CCGTATATCACCTCACTCAACTCAACCAACTCCACCAGCTCAGGAAGCATACTACGCCCA 240  
QY 3215 ACCTGCCACCGGTCTCACTAAAGTCTAGTAAACAACTGTAAACCAAGTAACTGACAAAAATCC 3274  
Db 241 ACCTGCCCTCCGCTCACTAAAGTCTTGTCAACCACTGTAAGCCAGTAACTGACAAAGATCC 300  
QY 3275 ACAAGCATATAAGAGCAACAAACAAATATAGGTTTGAATCATCGGAGAGAGAGAAATCG 3334  
|||||  
Db 301 ACAAGGCTATATAAGAGCAACAAACAAATACAGGTTTGAATCATCGGAGAGAGAGAAATCG 360  
QY 3335 CTTTTAAATGATTCGTACCAACGTTTCTCATGTGGTTGGCCAGTTCGATGACATAAGAA 3394  
|||||  
Db 361 CTTTCAGATATACGAACCAATGTTTCTCATGTGGTTGGCCAGTTCGATGACATCAGAA 420  
QY 3395 AAAACCCCTAGGAAGTCTTGTTCCTGAATGACAACTTGACCAACATCATATAAGATGCTC 3454  
|||||  
Db 421 AAAACCCCGAGGAGTCTGTTGTCTGAATGACAACTTGACCAACATCATATAAGATGCTC 480  
QY 3455 AGACAGTGAAGGCTGTTCTCAGGAGCTTCTATGAATCCATGTTCCCATACCTTCCCAAT 3514  
|||||  
Db 481 GGACAGTGAAGGCTGTTCTCAGGAGCTTCTATGATGCTCCATGTTTCCCATACCTTCCCAAT 540  
QY 3515 TTGAAGTCCCAAGAGAGTATCGAAACGTTTCTTCATATGTCATGACGATGACGAGAAATGGA 3574  
Db 541 TTGAAGTCCCAAGAGAGTATCGAAACGCTTCTGACATGTCATGACGATGACGAGAAATGGA 600  
QY 3575 GGGCTTATCGAGACAAATTTGAAAT 3598  
Db 601 GGGCATATCGAGACAACTGAAAT 624

Search completed: April 13, 2003, 01:34:36  
Job time : 3215 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 12, 2003, 22:47:35 ; Search time 6160 Seconds  
(without alignments)  
17008.133 Million cell updates/sec

Title: US-10-023-888-1

Perfect score: 3600

Sequence: 1 atggagacagacacactctct.....atcgagacaaattgaagtag 3600

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb.ba.\*

2: gb.htg.\*

3: gb.in.\*

4: gb.ov.\*

5: gb.pat.\*

6: gb.ph.\*

7: gb.pl.\*

8: gb.pr.\*

9: gb.ro.\*

10: gb.sts.\*

11: gb.sy.\*

12: gb.un.\*

13: gb.vi.\*

14: em.ba.\*

15: em.fun.\*

16: em.in.\*

17: em.ju.\*

18: em.or.\*

19: em.ov.\*

20: em.pat.\*

21: em.ph.\*

22: em.pl.\*

23: em.ro.\*

24: em.sts.\*

25: em.un.\*

26: em.vi.\*

27: em.htg\_hum.\*

28: em.htg\_inv.\*

29: em.htg\_other.\*

30: em.htg\_mus.\*

31: em.htg\_pln.\*

32: em.htg\_rod.\*

33: em.htg\_mam.\*

34: em.htg\_vrt.\*

35: em.sv.\*

36: em.htgo\_hum.\*

37: em.htgo\_mus.\*

38: em.htgo\_other.\*

39: em.htgo\_hum.\*

40: em.htgo\_mus.\*

41: em.htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	3472	96.4	4369	6	AX468102	Sequence
2	3100.4	86.1	4277	6	AX468106	Sequence
3	2685.4	74.6	4511	9	AB033034	Homo sapi
4	2404.8	66.8	2603	9	AK056137	Homo sapi
5	1103.6	30.7	162427	9	AC010205	Homo sapi
6	1103.6	30.7	174231	9	AC063950	Homo sapi
7	1103.2	30.6	177364	9	AC005409	Homo sapi
8	849.8	23.6	1846	10	MUSBDLZ	L36434 Mus Musculu
9	592.8	16.5	2428	9	HS0802690	AL3359588 Homo sapi
10	499	13.9	141215	2	AC121402	AC121402 Rattus no
11	462.4	12.8	281892	2	AC125486	AC125486 Mus muscu
12	403.8	11.2	2332	9	HS0803479	AL832172 Homo sapi
13	266	7.4	2076	9	AK001821	AK001821 Homo sapi
14	214.4	6.0	142667	10	AL135758	AL135758 Mouse DNA
15	184.8	5.1	12996	10	AC005501	AC005501 Mus muscu
16	183.4	5.1	198967	2	AC124926	AC124926 Rattus no
17	165.6	4.6	281892	2	AC125486	AC125486 Mus muscu
18	157.6	4.4	141215	2	AC121402	AC121402 Rattus no
19	157.6	4.4	177607	2	AC110966	AC110966 Rattus no
20	157	4.4	177607	2	AC110966	AC110966 Rattus no
21	138	3.8	2610	3	AY089618	AY089618 Drosophil
22	138	3.8	51430	2	AC017981	AC017981 Drosophil
23	138	3.8	189757	3	AC008186	AC008186 Drosophil
24	138	3.8	243313	3	AE003834	AE003834 Drosophil
25	119.4	3.3	46555	2	AC100621	AC100621 Mus muscu
26	102.6	2.8	42816	1	SCC88	AL139298 Streptomy
27	100.4	2.8	36849	1	SC1C3	AL023702 Streptomy
28	91.4	2.5	15347	6	AX283713	AX283713 Sequence
29	91.4	2.5	28170	6	AX283721	AX283721 Sequence
30	75.4	2.1	14624	1	SSA17900	Y17900 Streptococc
31	75.4	2.1	20555	6	AX283717	AX283717 Sequence
32	74.4	2.1	18274	6	AX283720	AX283720 Sequence
33	73	2.0	340806	1	NNA12491	AL162752 Neisseria
34	72	2.0	780	9	BC002779	BC002779 Homo sapi
35	71.4	2.0	5064	1	AF019760	AF019760 Neisseria
36	69.6	1.9	4798	1	AF518558	AF518558 Actinobac
37	67	1.9	7218	6	I66494	I66494 Sequence 14
38	66.6	1.8	5070	6	AX234391	AX234391 Sequence
39	65.6	1.8	1779	6	AX451462	AX451462 Sequence
40	65.6	1.8	2670	6	E59779	E59779 Soluble neu
41	65.2	1.8	624	6	AX084406	AX084406 Sequence
42	65.2	1.8	624	6	AX128409	AX128409 Sequence
43	65.2	1.8	5732	1	AF112478	AF112478 Neisseria
44	64	1.8	5732	6	AX027904	AX027904 Sequence
45	64	1.8	5732	12	AX027910	AX027910 Sequence

# ALIGNMENTS

RESULT 1	AX468102	Sequence 25 from Patent WO0250279.	4369 bp	DNA	linear	PAT 16-JUL-2002
LOCUS	AX468102	Sequence 25 from Patent WO0250279.				
DEFINITION	AX468102	Sequence 25 from Patent WO0250279.				
ACCESSION	AX468102.1	GI:21900976				
VERSION	AX468102.1	GI:21900976				
KEYWORDS	human.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;					
AUTHORS	Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.					
	Baughin, M.R., Lu, Y., Arvizu, C., Ramkumar, J., Yao, M.G.,					
	Policky, J.L., Walia, N.K., Tribouley, K.M., Yue, H., Batra, S.,					
	Ding, L., Lal, P.G., Borowsky, M.L., Lu, D.A., Gandhi, A.R.,					

Griffin, J.A., Xu, Y., Azimzai, Y., Gietzen, K.J., Tang, Y.T.,  
Warren, B.A., Mason, P.M., Burford, N.M., Hafalia, A.J., Lee, E.A.,  
Yang, J., Gorvad, A.E., Emerling, B.M., Marquis, J.P., Lee, S.Y.,  
Swanekar, A. and Reddy, R.  
Nucleic acid-associated proteins  
Patent: WO 0250279-A 25 27-JUN-2002;  
Incyte Genomics, Inc. (US)  
Location/Qualifiers  
source  
1. 4369  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/note="Incyte ID No: 2749402CB1"  
BASE COUNT 1339 a 912 c 978 g 1140 t  
ORIGIN

Query Match 96.4%; Score 3472; DB 6; Length 4369;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 3481; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 103 AGCCGAGATCAATACCATGTTTGTGATTCCTATAGAGACAATATGCTGGAAAGTCC 162  
|||||  
Db 266 AGCCGAGATCAATACCATGTTTGTGATTCCTATAGAGACAATATGCTGGAAAGTCC 325  
|||||

QY 163 TTTCAGATCGCTTGTGCTGCCATGCCGATGACGTTGTTACACCTGGGTGAATGGC 222  
|||||  
Db 326 TTTCAGATCGCTTGTGCTGCCATGCCGATGACGTTGTTACACCTGGGTGAATGGC 385  
|||||

QY 223 ACAGATCTTGAACACTACTGAAGAACTACAGCAGGTCAGAGAACAGATGAGAGAGCAG 282  
|||||  
Db 386 ACAGATCTTGAACACTACTGAAGAACTACAGCAGGTCAGAGAACAGATGAGAGAGCAG 445  
|||||

QY 283 AAAGCAATGAGAGAAATCCTTGGGAAAAACACACGGAACCTACTAAGAGAGTGAAG 342  
|||||  
Db 446 AAAGCAATGAGAGAAATCCTTGGGAAAAACACACGGAACCTACTAAGAGAGTGAAG 505  
|||||

QY 343 CAGTTAGAGCTTTGCTTAACACACTGCAATTAAGTGCCAAATGCTTGTCCCTGGACCCAGCC 402  
|||||  
Db 506 CAGTTAGAGCTTTGCTTAACACACTGCAATTAAGTGCCAAATGCTTGTCCCTGGACCCAGCC 565  
|||||

QY 403 CTGCCAGCAACATCACCTCTGAAGGACCTGCCATCTCTTTATCCTCTCTTTTCAATCTGCC 462  
|||||  
Db 566 CTGCCAGCAACATCACCTCTGAAGGACCTGCCATCTCTTTATCCTCTTTTCAATCTGCC 625  
|||||

QY 463 AGTGACATTTCAATGTTGGAACCAACCAAAACCCCTCTACCAATGCTCAGTTGTTGT 522  
|||||  
Db 626 AGTGACATTTCAATGTTGGAACCAACCAAAACCCCTCTACCAATGCTCAGTTGTTGT 685  
|||||

QY 523 TTTCAGTACTTAAGGATGTTGAAGATGCCACCTGAGCTGCTTAAAGGAAATAGCAGA 582  
|||||  
Db 686 TTTCAGTACTTAAGGATGTTGAAGATGCCACCTGAGCTGCTTAAAGGAAATAGCAGA 745  
|||||

QY 583 CAGACAGTATGAGGGGCTACTTGACACAGATAAAGAGTCCCTGGATTAGTGTCTAATG 642  
|||||  
Db 746 CAGACAGTATGAGGGGCTACTTGACACAGATAAAGAGTCCCTGGATTAGTGTCTAATG 805  
|||||

QY 643 CAAGATTTGGCTTCTCCTGAGTGGATTTCCACCAACATFCAAGGAAACAAATCAACTAAA 702  
|||||  
Db 806 CAAGATTTGGCTTCTCCTGAGTGGATTTCCACCAACATFCAAGGAAACAAATCAACTAAA 865  
|||||

QY 703 ACAGATTTGCCAGAAAATCTTTCTCTTAAAGTCAAACTGTTGACGTTGATTCAGAGGCC 762  
|||||  
Db 866 ACAGATTTGCCAGAAAATCTTTCTCTTAAAGTCAAACTGTTGACGTTGATTCAGAGGCC 925  
|||||

QY 763 AGTGTAGGCTTCTAAAACCTGAATAACCCCAAGGATTTTCAAGAAATGAATAAGCAAACT 822  
|||||  
Db 926 AGTGTAGGCTTCTAAAACCTGAATAACCCCAAGGATTTTCAAGAAATGAATAAGCAAACT 985  
|||||

QY 823 AAGAAGACATGACCATTTGATGGAAGAACTGACCAATAGTCTGATATTTATATGG 882  
|||||  
Db 986 AAGAAGACATGACCATTTGATGGAAGAACTGACCAATAGTCTGATATTTATATGG 1045  
|||||

QY 883 GATCTGAGCGCCATCAGCGCTCTAAGCAGGATGAAGACATCTCTGCCAGTCTGTTTGA 942  
|||||

Db 1046 GATCTGAGCGCCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCCAGTCTGTTTGA 1105  
QY 943 GATAACGAAGAACTAGGTACTCATTCGATCTATCGAGAGCATGCACCATGGTTCGG 1002  
|||||  
Db 1106 GATAACGAAGAACTAGGTACTCATTCGATCTATCGAGAGCATGCACCATGGTTCGG 1165  
|||||

QY 1003 AATATTTTTCATTTGTCACCAACGGGAGATTCATCTGCTGAACTTGACAATCTCGA 1062  
|||||  
Db 1166 AATATTTTTCATTTGTCACCAACGGGAGATTCATCTGCTGAACTTGACAATCTCGA 1225  
|||||

QY 1063 GTGACAAATAGTAACACACAGGATGTTTTCGAAAATTTGAGCCACTTGCTACCTTATG 1122  
|||||  
Db 1226 GTGACAAATAGTAACACACAGGATGTTTTCGAAAATTTGAGCCACTTGCTACCTTATG 1285  
|||||

QY 1123 TCACCTGCTATTGAAAAGTCACTCATCGCATCGAAGGCTGTCCAGAAAGTTTATTTAC 1182  
|||||  
Db 1286 TCACCTGCTATTGAAAAGTCACTCATCGCATCGAAGGCTGTCCAGAAAGTTTATTTAC 1345  
|||||

QY 1183 CTAATATGATGATGTCATGTTTGGGAAAGATGTCGTGCCAGATGATTTTACAGTCACTCC 1242  
|||||  
Db 1346 CTAATATGATGATGTCATGTTTGGGAAAGATGTCGTGCCAGATGATTTTACAGTCACTCC 1405  
|||||

QY 1243 AAAGGCCAGAAAGTATTGTCATGCGCTGTGCCAACTGTGCCGAGGCTGCCAGT 1302  
|||||  
Db 1406 AAAGGCCAGAAAGTATTGTCATGCGCTGTGCCAACTGTGCCGAGGCTGCCAGT 1465  
|||||

QY 1303 TCCTGGATTAAAGATGCTATTGTCAGAAAGCTTGTAAATTCAGCCCTGCGATTGGAT 1362  
|||||  
Db 1466 TCCTGGATTAAAGATGCTATTGTCAGAAAGCTTGTAAATTCAGCCCTGCGATTGGAT 1525  
|||||

QY 1363 GGTGGGATTTGCTCGGAAACAGTGGAGGAGTTCGCTATATTGCGAGGAGTGGAGTACT 1422  
|||||  
Db 1526 GGTGGGATTTGCTCGGAAACAGTGGAGGAGTTCGCTATATTGCGAGGAGTGGAGTACT 1585  
|||||

QY 1423 GGAGTATTGGAGTGGACAGCCCTGGCAGTTTGGTGGAGGAATAAACAGTCTCTTAC 1482  
|||||  
Db 1586 GGAGTATTGGAGTGGACAGCCCTGGCAGTTTGGTGGAGGAATAAACAGTCTCTTAC 1645  
|||||

QY 1483 TGTAATCAGGATGTCGGAATTCCTGGCTCGCTGATTAAGTTCTGTGACCAAGCATGCAAT 1542  
|||||  
Db 1646 TGTAATCAGGATGTCGGAATTCCTGGCTCGCTGATTAAGTTCTGTGACCAAGCATGCAAT 1705  
|||||

QY 1543 GTCTTGTCTGTGGGTTTGTATGCTGCGGACTGTGGGCAAGATCATTTTCATGAATGTAT 1602  
|||||  
Db 1706 GTCTTGTCTGTGGGTTTGTATGCTGCGGACTGTGGGCAAGATCATTTTCATGAATGTAT 1765  
|||||

QY 1603 AAAGTGTATCTCTCCCAACACAGACTCACTATATTATTTCCAAAAGGTGAATGCGCTGCC 1662  
|||||  
Db 1766 AAAGTGTATCTCTCCCAACACAGACTCACTATATTATTTCCAAAAGGTGAATGCGCTGCC 1825  
|||||

QY 1663 TATTTTCAGCTTTGCGAAGTAGCCAAAAGAGGAGTTGAAGGTGCTTATAGTGACAATCCA 1722  
|||||  
Db 1826 TATTTTCAGCTTTGCGAAGTAGCCAAAAGAGGAGTTGAAGGTGCTTATAGTGACAATCCA 1885  
|||||

QY 1723 ATAATTCGACATGCTTCTATTGCGCAACAGTGGAAACCATCCACCTCATATATGCACAT 1782  
|||||  
Db 1886 ATAATTCGACATGCTTCTATTGCGCAACAGTGGAAACCATCCACCTCATATATGCACAT 1945  
|||||

QY 1783 GGAATGAATGCCACCAATACATTTTATCTCACGTTTCAAAATACAAACCATGAAGAG 1842  
|||||  
Db 1946 GGAATGAATGCCACCAATACATTTTATCTCACGTTTCAAAATACAAACCATGAAGAG 2005  
|||||

QY 1843 TTCAAAATCGAGATAACAGTGGAGTGGACACAAAGGAGGACCAAACTGAATTTACG 1902  
|||||  
Db 2006 TTCAAAATCGAGATAACAGTGGAGTGGACACAAAGGAGGACCAAACTGAATTTACG 2065  
|||||

QY 1903 GCCCAGAGGGTTACGAAAATTTAGTTAGTCCCATTAACACTTCTTCCAGAGGCGGAATC 1962  
|||||  
Db 2066 GCCCAGAGGGTTACGAAAATTTAGTTAGTCCCATTAACACTTCTTCCAGAGGCGGAATC 2125  
|||||

QY 1963 CTTTTCAGGATATTCGAAAAGAAAACCCCTTCCGGAAGTTTAAAGAGACATGATGTTTAC 2022  
|||||  
Db 2126 CTTTTCAGGATATTCGAAAAGAAAACCCCTTCCGGAAGTTTAAAGAGACATGATGTTTAC 2185  
|||||





Db 273 AGCCAGATCAATACCATGTTTGTGATTCCTATAGACAAATATGCTGGAAATCC 332  
Qy 163 TTTCCAGATCGGCTTTGTCTGCCATGCCGATTCGACCTGTGTTACACCTGGGTGAATGGC 222  
Db 333 TTTCCAGATCGGCTTTGTCTGCCATGCCGATTCGACCTGTGTTACACCTGGGTGAATGGC 392  
Qy 223 ACAGATCTTGAACACTACGAGAACTACAGCAGGTTCAGAACACAGATGGAGGAGCAG 282  
Db 393 ACAGATCTTGAACACTACGAGAACTACAGCAGGTTCAGAACACAGATGGAGGAGCAG 452  
Qy 283 AAAGCAATGAGAGAAATCCTTGGGAAACACACACGAACTACTAAGAGAGTGAGAAG 342  
Db 453 AAAGCAATGAGAGAAATCCTTGGGAAACACACACGAACTACTAAGAGAGTGAGAAG 512  
Qy 343 CAGTTAGAGTCTTGTCTAAACACACTGCATTAAGGTGCAATGCTTGTCTGGACCCAGCC 402  
Db 513 CAGTTAGAGTCTTGTCTAAACACACTGCATTAAGGTGCAATGCTTGTCTGGACCCAGCC 572  
Qy 403 CTGCCAGCAACATACACCTGGAAGACCTGCGCACTCTTTATCCTTTTTCATCTGTC 462  
Db 573 CTGCCAGCAACATACACCTGGAAGACCTGCGCACTCTTTATCCTTTTTCATCTGTC 632  
Qy 463 AGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGCTCAGTTGTTGTT 522  
Db 633 AGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGCTCAGTTGTTGTT 692  
Qy 523 TTTGACAGTACTAAGGATGTTGAAGATGCCACCTCTGAGCTGCTTAAGGAAATAGCAGA 582  
Db 693 TTTGACAGTACTAAGGATGTTGAAGATGCCACCTCTGAGCTGCTTAAGGAAATAGCAGA 752  
Qy 583 CAGACAGTATGGAGGGCTACTTGACACACAGATAAAGTCCCTGAGTATGCTGCTAATG 642  
Db 753 CAGACAGTATGGAGGGCTACTTGACACACAGATAAAGTCCCTGAGTATGCTGCTAATG 812  
Qy 643 CAAGATTTGGCTTTCTCCTGAGTGGATTTCCACCACATTCAGGAAACAAATCAACTAAA 702  
Db 813 CAAGATTTGGCTTTCTCCTGAGTGGATTTCCACCACATTCAGGAAACAAATCAACTAAA 872  
Qy 703 ACAAAATGGCCAGAAATCTTCTCTAAAGTCAAACTGTTGCAAGTGTGTTATCAGAGGCC 762  
Db 873 ACAAAATGGCCAGAAATCTTCTCTAAAGTCAAACTGTTGCAAGTGTGTTATCAGAGGCC 932  
Qy 763 AGTGAGGCTTTCTAAACACTGAATAACCCCAAGGATTTCAAGAAATGAATAAGCAAACT 822  
Db 933 AGTGAGGCTTTCTAAACACTGAATAACCCCAAGGATTTCAAGAAATGAATAAGCAAACT 992  
Qy 823 AAGAGACATGACCATGATGAGGAAAGAACTGACCAATAGTCTGCAATATTTATATGG 882  
Db 993 AAGAGAACATGACCATGATGAGGAAAGAACTGACCAATAGTCTGCAATATTTATATGG 1052  
Qy 883 GATCTGAGCGCATCAGCAGTCTAAGCAGATGAAGACATCTCTGCGAGTCTGTTTGA 942  
Db 1053 GATCTGAGCGCATCAGCAGTCTAAGCAGATGAAGACATCTCTGCGAGTCTGTTTGA 1112  
Qy 943 GATAACGAAGAACTGAGGTACTCTATCCGATCTATCGAGAGGATGACCATGGGTTCGG 1002  
Db 1113 GATAACGAAGAACTGAGGTACTCTATCCGATCTATCGAGAGGATGACCATGGGTTCGG 1172  
Qy 1003 AATATTTTCATTTGTCACCAACGGGCAGATTCATCTGCTGGCTGACACATCCTCGA 1062  
Db 1173 AATATTTTCATTTGTCACCAACGGGCAGATTCATCTGCTGGCTGACACATCCTCGA 1232  
Qy 1063 GTGACATAGTAACACACAGGATGTTTTCGAAATTTGAGCCACTGCGCTACCTTAGT 1122  
Db 1233 GTGACATAGTAACACACAGGATGTTTTCGAAATTTGAGCCACTGCGCTACCTTAGT 1292  
Qy 1123 TCACCTGCTATTTGAAATCAGTTCATCGCATCGAAGGGCTGTCCAGAAAGTTTATTTAC 1182  
Db 1293 TCACCTGCTATTTGAAATCAGTTCATCGCATCGAAGGGCTGTCCAGAAAGTTTATTTAC 1352  
Qy 1183 CTAATGATGATGTCATGTTGGGAAGGATGCTGCGCCAGATGATTTTACATGCTACTCC 1242  
Db 1353 CTAATGATGATGTCATGTTGGGAAGGATGCTGCGCCAGATGATTTTACATGCTACTCC 1412

Qy 1243 AAAGGCCAGAGGTTTATTTGACATGGCTGTGCCAAACTGTGCCGAGGCTGCCAGGT 1302  
Db 1413 AAAGGCCAGAGGTTTATTTGACATGGCT----- 1442  
Qy 1303 TCTCGATTAAAGATGGCTATTGTGACAAGGCTTTGTAATAATTCAGCCTGGATTTGGGAT 1362  
Db 1443 ----- 1442  
Qy 1363 GGTGGGATTGCTCTGGAACAGTGGAGGAGTCCCTATATTGACGAGGTGGAGGTACT 1422  
Db 1443 ----- 1442  
Qy 1423 GGGAGTATTGGAGTTGGACAGCCCTGGCAGTTTGGTGGAGGAATAAACAGTGTCTCTTAC 1482  
Db 1443 -----TTTGGTGGAGGAATAAACAGTGTCTCTTAC 1472  
Qy 1483 TGTAAATCAGGATGTCGCAATTCCTGGCTGCTGATATAAGTCTGTGACCAAGCATGCAAT 1542  
Db 1473 TGTAAATCAGGATGTCGCAATTCCTGGCTGCTGATATAAGTCTGTGACCAAGCATGCAAT 1532  
Qy 1543 GTCTTGTCTGTGGTGTGATGCTGCGACTGTGGCAAGATCAATTTTCATGAATTTGAT 1602  
Db 1533 GTCTTGTCTGTGGTGTGATGCTGCGACTGTGGCAAGATCAATTTTCATGAATTTGAT 1592  
Qy 1603 AAAGTGATCTTCTCCCAAAACAGACTCAGTATATTTTCCAAAAGGTGAATGCCTGCCT 1662  
Db 1593 AAAGTGATCTTCTCCCAAAACAGACTCAGTATATTTTCCAAAAGGTGAATGCCTGCCT 1652  
Qy 1663 TATTTTCAGCTTTGCAAGTAGTCCCAAAAGAGAGGATTTGAAGTGCCTATAGTGAACAATCCA 1722  
Db 1653 TATTTTCAGCTTTGCAAGTAGTCCCAAAAGAGAGGATTTGAAGTGCCTATAGTGAACAATCCA 1712  
Qy 1723 ATAAATCGACATGCTTCTATTGCCAACAGTGGAAACCATCCCTCATATGACACAGT 1782  
Db 1713 ATAAATCGACATGCTTCTATTGCCAACAGTGGAAACCATCCCTCATATGACACAGT 1772  
Qy 1783 GGAATGAATGCCACACACATACATTTTAACTCTACGTTTCAAAATACAAACGATGAAGAG 1842  
Db 1773 GGAATGAATGCCACACACATACATTTTAACTCTACGTTTCAAAATACAAACGATGAAGAG 1832  
Qy 1843 TTCAAATCGAGATAACAGTGGAGTGACACAAAGGGAGGACCAAACTGAATTTCTACG 1902  
Db 1833 TTCAAATCGAGATAACAGTGGAGTGACACAAAGGGAGGACCAAACTGAATTTCTACA 1892  
Qy 1903 GCCCAGAGGGTTACGAAATTTAGTTAGTCCCAATAACACTTCTTCCAGAGCGGAAATC 1962  
Db 1893 GCCCAGAGGGTTACGAAATTTAGTTAGTCCCAATAACACTTCTTCCAGAGCGGAAATC 1952  
Qy 1963 CTTTGTGAGGATATTCCTCAAGAAACGCTTCCGAAAGTTTAAAGACATGATGTTAAC 2022  
Db 1953 CTTTGTGAGGATATTCCTCAAGAAACGCTTCCGAAAGTTTAAAGACATGATGTTAAC 2012  
Qy 2023 TCAACAGAGAGAGCCAGAGAGGTGAAATTTCCCTCGTAAATTTTCACTCTCTTCCA 2082  
Db 2013 TCAACAGAGAGAGCCAGAGAGGTGAAATTTCCCTCGTAAATTTTCACTCTCTTCCA 2072  
Qy 2083 AAAGAGCCCGAGTTGAGTCTCAATACCTTTGATTTGCAACTGGAAACATGGAGACATCACT 2142  
Db 2073 AAAGAGCCCGAGTTGAGTCTCAATACCTTTGATTTGCAACTGGAAACATGGAGACATCACT 2132  
Qy 2143 TTGAAGGATACAAATTTGTCCAAGTCAAGCTTGTGCTGAGATCATTTTCTGATGAACCTACAG 2202  
Db 2133 TTGAAGGATACAAATTTGTCCAAGTCAAGCTTGTGCTGAGATCATTTTCTGATGAACCTACAG 2192  
Qy 2203 CATGCTAAAAATAAAAAATCAAGCTATAATAACAGATGAACAAATGACAGTTTGGTGGCT 2262  
Db 2193 CATGCTAAAAATAAAAAATCAAGCTATAATAACAGATGAACAAATGACAGTTTGGTGGCT 2252  
Qy 2263 CCACAGGAAAAACAGGTTTCATAAAGATCTTTGCCAAACAGCTTAGGAGTGTCTGAAAGA 2322  
Db 2253 CCACAGGAAAAACAGGTTTCATAAAGATCTTTGCCAAACAGCTTAGGAGTGTCTGAAAGA 2312





BASE COUNT 1382 a 898 c 938 g 1293 t

Query Match 74.63; Score 2685.4; DB 9; Length 4511;  
Best Local Similarity 99.4%; Pred. No. 0;  
Matches 2695; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY	888	GAGCGCCATCAGCCAGTCTTAAGCAGGATGAAGACATCTCGCCAGTCGTTTGAAGATAA	947
Db	1	GAGCGCCATCAGCCAGTCTTAAGCAGGATGAAGACATCTCGCCAGTCGTTTGAAGATAA	60
QY	948	CGAAGAACTGAGGTACTCATTTGCGATCTATCGAGAGGCATGCACCATGGGTTTCGGAATAT	1007
Db	61	CGAAGAACTGAGGTACTCATTTGCGATCTATCGAGAGGCATGCACCATGGGTTTCGGAATAT	120
QY	1008	TTTTCATTGTCACCAACGGCGAGATTCACCTCGCTGCGTGAACCTTGACAATCCTCGAGTGAC	1067
Db	121	TTTTCATTGTCACCAACGGCGAGATTCACCTCGCTGCGTGAACCTTGACAATCCTCGAGTGAC	180
QY	1068	AATAGTAAACACACCAAGGATGTTTTCGAAATTTTGAGCCACTTGCCTACCTTTAGTTTCACC	1127
Db	181	AATAGTAAACACACCAAGGATGTTTTCGAAATTTTGAGCCACTTGCCTACCTTTAGTTTCACC	240
QY	1128	TGCTATTGAAAGTCACGTTTCATCCATCGAAGGGCTGTGCCAGAAGTTTATTACCTAAA	1187
Db	241	TGCTATTGAAAGTCACATTCATCCATCGAAGGGCTGTGCCAGAAGTTTATTACCTAAA	300
QY	1188	TGATGATGTCATGTTTGGGAAGGATGCTGGCCAGATGATTTTACAGTCACTCCCAAAGG	1247
Db	301	TGATGATGTCATGTTTGGGAAGGATGCTGGCCAGATGATTTTACAGTCACTCCCAAAGG	360
QY	1248	CCAGAAGGTTTATTTGACATGGCCCTGTGCCAAACTGTGCCGAGGGTGCCCAAGTTCCTG	1307
Db	361	CCAGAAGGTTTATTTGACATGGCCCTGTGCCAAACTGTGCCGAGGGTGCCCAAGTTCCTG	420
QY	1308	GATTAAAGGATGGCTATTGTGACAGGCTTGTAATAATCAGCCTCGGATGGGATGGTG	1367
Db	421	GATTAAAGGATGGCTATTGTGACAGGCTTGTAATAATCAGCCTCGGATGGGATGGTG	480
QY	1368	GGATTGCTCTGGAACAGTGAGGAGTGCCTATATTGAGGAGGTGGAGTACTGGGAG	1427
Db	481	GGATTGCTCTGGAACAGTGAGGAGTGCCTATATTGAGGAGGTGGAGTACTGGGAG	540
QY	1428	TATTTGGAGTTGGACAGCCCTGGCAGTTTGGTGGGAAATAACAGTGTCTTACTGTAA	1487
Db	541	TATTTGGAGTTGGACAGCCCTGGCAGTTTGGTGGGAAATAACAGTGTCTTACTGTAA	600
QY	1488	TCAGGAGTTGCGAATTCCTGGCTGCGTGATAAGTTCTGTGACCAAGCATGCAATGTCTT	1547
Db	601	TCAGGAGTTGCGAATTCCTGGCTGCGTGATAAGTTCTGTGACCAAGCATGCAATGTCTT	660
QY	1548	GTCTGTGGGTTTGATGTGCGGACTGTGGCAAGATCATTTTTCATGAATGTATAAAAGT	1607
Db	661	GTCTGTGGGTTTGATGTGCGGACTGTGGCAAGATCATTTTTCATGAATGTATAAAAGT	720
QY	1608	GATCTCTTCCCAACACAGACTCACTATATATTCCAAAGGTGAATGCGCTGCCCTATT	1667
Db	721	GATCTCTTCCCAACACAGACTCACTATATATTCCAAAGGTGAATGCGCTGCCCTATT	780
QY	1668	CAGCTTTTCGAGAAGTAGCCAAAGAGAGGTTGAAGTGCGCTATAGTGACAATCCAAATAT	1727
Db	781	CAGCTTTTCGAGAAGTAGCCAAAGAGAGGTTGAAGTGCGCTATAGTGACAATCCAAATAT	840
QY	1728	TCGACATGCTTCTATTGCCAACAGTGGAAACCAATCCACCTCATTAATGCACAGTGGAA	1787



QY 343 CAGTTAGAGTGTCTTAACACACTGCATTAAGGTGCCAATGCTTGCTCGACCCAGCC 402  
Db |||||  
QY 436 CAGTTAGAGTGTCTTAACACACTGCATTAAGGTGCCAATGCTTGCTCGACCCAGCC 495  
Db |||||  
QY 403 CTGCCAGCCACATCACCCTCAAGGACCTGCATCTCTTTATCCCTCTTTTCATCTCGCC 462  
Db |||||  
QY 496 CTGCCAGCCACATCACCCTCAAGGACCTGCATCTCTTTATCCCTCTTTTCATCTCGCC 555  
QY 463 AGTGACATTTCAATGTGTGAAAACCAAAAACCCCTCTACCAATGTCTCAGTTGTGTT 522  
Db |||||  
QY 556 AGTGACATTTCAATGTGTGAAAACCAAAAACCCCTCTACCAATGTCTCAGTTGTGTT 615  
QY 523 TTTGACAGTACTAAGGATGTGAAGATGCCACACTCTCGACTCTCTTAAGGAAATAGCAGA 582  
Db |||||  
QY 616 TTTGACAGTACTAAGGATGTGAAGATGCCACACTCTCGACTCTCTTAAGGAAATAGCAGA 675  
QY 583 CAGACAGTATGGAGGGCTACTTGGACACAGATAAAGAAAGTCCCTGGATTAAGTCTAATG 642  
Db |||||  
QY 676 CAGACAGTATGGAGGGCTACTTGGACACAGATAAAGAAAGTCCCTGGATTAAGTCTAATG 735  
QY 643 CAAGATTTGGCTTCTTCCCTGAGTGATTTCCACCAACATTCAGGAAACAATCAACTAAAA 702  
Db |||||  
QY 736 CAAGATTTGGCTTCTTCCCTGAGTGATTTCCACCAACATTCAGGAAACAATCAACTAAAA 795  
QY 703 ACAAATTTGCCAGAAAATCTTCTCTTAAGTCAAACTGTGCAGTTGTATTCAGAGGC 762  
Db |||||  
QY 796 ACAAATTTGCCAGAAAATCTTCTCTTAAGTCAAACTGTGCAGTTGTATTCAGAGGC 855  
QY 763 AGTGACGGCTTCTAAACTGATTAACCCCAAGGATTTCAAGAAATTAAGCAAACT 822  
Db |||||  
QY 856 AGTGACGGCTTCTAAACTGATTAACCCCAAGGATTTCAAGAAATTAAGCAAACT 915  
QY 823 AAGAAGAACATGACCAATTTGATGAAAGAACTGACCATTAAGTCTGATTTATATGG 882  
Db |||||  
QY 916 AAGAAGAACATGACCAATTTGATGAAAGAACTGACCATTAAGTCTGATTTATATGG 975  
QY 883 GATCTGAGGCCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCCAGTCTGTTTGA 942  
Db |||||  
QY 976 GATCTGAGGCCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCCAGTCTGTTTGA 1035  
QY 943 GATAACGAAGAACTGAGTACTCATTTGCGATCTATCGAGAGGATGACCATGGTTCGG 1002  
Db |||||  
QY 1036 GATAACGAAGAACTGAGTACTCATTTGCGATCTATCGAGAGGATGACCATGGTTCGG 1095  
QY 1003 AATATTTTCATTTGTCACCAACGGGAGATTCATCTCTGCTGAACTTTGACAACTCTCGA 1062  
Db |||||  
QY 1096 AATATTTTCATTTGTCACCAACGGGAGATTCATCTCTGCTGAACTTTGACAACTCTCGA 1155  
QY 1063 GTGACAAATAGTAAACACACAGGATGTTTTCGAAATTTGAGCCACTTGCTACCTTTAGT 1122  
Db |||||  
QY 1156 GTGACAAATAGTAAACACACAGGATGTTTTCGAAATTTGAGCCACTTGCTACCTTTAGT 1215  
QY 1123 TCACCTGCTATTGAAGTCACTGTCATCGATCGAAGGGCTGTCCAGAAAGTTTATTTC 1182  
Db |||||  
QY 1216 TCACCTGCTATTGAAGTCACTGTCATCGATCGAAGGGCTGTCCAGAAAGTTTATTTC 1275  
QY 1183 CTAAATGATGATGTCATGTTTGGGAAGGATGCTGCGCAGATGATTTTACAGTCACTCC 1242  
Db |||||  
QY 1276 CTAAATGATGATGTCATGTTTGGGAAGGATGCTGCGCAGATGATTTTACAGTCACTCC 1335  
QY 1243 AAAGGCCAGAAAGTTTATTGACATGGCTGTGCCAAACTGTGCCAGGGCTGCCAGGT 1302  
Db |||||  
QY 1336 AAAGGCCAGAAAGTTTATTGACATGGCTGTGCCAAACTGTGCCAGGGCTGCCAGGT 1395  
QY 1303 TCCTGGATTAAGGATGGCTATTGTGACAAAGCTTGTATAAATTCAGCTGCGATTGGGAT 1362  
Db |||||  
QY 1396 TCCTGGATTAAGGATGGCTATTGTGACAAAGCTTGTATAAATTCAGCTGCGATTGGGAT 1455  
QY 1363 GGTGGGGATGCTCTGGAACACAGTGGAGGGATGCGTATATTGAGAGGTGGAGGTACT 1422  
Db |||||  
QY 1456 GGTGGGGATGCTCTGGAACACAGTGGAGGGATGCGTATATTGAGAGGTGGAGGTACT 1515

QY 1423 GGGAGTATTGGAGTTGGACAGCCCTGGCAGTTTGGTGGAGGAATAAACAGTGTCTCTTAC 1482  
Db |||||  
QY 1516 GGGAGTATTGGAGTTGGACAGCCCTGGCAGTTTGGTGGAGGAATAAACAGTGTCTCTTAC 1575  
Db |||||  
QY 1483 TGTAAATCAGGGATGTCGGAATTCCTGGCTCGCTGATAAGTTCTGTGACCAAGCATGAAT 1542  
Db |||||  
QY 1576 TGTAAATCAGGGATGTCGGAATTCCTGGCTCGCTGATAAGTTCTGTGACCAAGCATGAAT 1635  
QY 1543 GTCTTCTCTGTGGGTTTGTGATGCTGGCAGTGTGGCAGAGATCATTTTCATGAATGTAT 1602  
Db |||||  
QY 1636 GTCTTCTCTGTGGGTTTGTGATGCTGGCAGTGTGGCAGAGATCATTTTCATGAATGTAT 1695  
QY 1603 AAAGTATCCTTCTCCCAACACAGACTCACTATATTTATTTCCAAAAGTGAATGCTGCT 1662  
Db |||||  
QY 1696 AAAGTATCCTTCTCCCAACACAGACTCACTATATTTATTTCCAAAAGTGAATGCTGCT 1755  
QY 1663 TATTTTCAGCTTTGCGAAGTAGCCAAAAGAGAGTTGAAGTGCCTATAGTGACAATCCA 1722  
Db |||||  
QY 1756 TATTTTCAGCTTTGCGAAGTAGCCAAAAGAGAGTTGAAGTGCCTATAGTGACAATCCA 1815  
QY 1723 ATAATTCGACATGCTTCTATTGCCCACCAAGTGGAAAACCATCCACCTCATATGCACAGT 1782  
Db |||||  
QY 1816 ATAATTCGACATGCTTCTATTGCCCACCAAGTGGAAAACCATCCACCTCATATGCACAGT 1875  
QY 1783 GGAATGAATGCCACACAAATACATTTTAACTCTCAGTTTCAAAAATACAAACGATGAAG 1842  
Db |||||  
QY 1876 GGAATGAATGCCACACAAATACATTTTAACTCTCAGTTTCAAAAATACAAACGATGAAG 1935  
QY 1843 TTCAAAATGCAGATAACAGTGGAGTGGACACAAAGGAGGACCAAACTGAATTCACG 1902  
Db |||||  
QY 1936 TTCAAAATGCAGATAACAGTGGAGTGGACACAAAGGAGGACCAAACTGAATTCACG 1995  
QY 1903 GCCAAGAGGGTTACGAAAATTTAGTTAGTCCCATACACATTTCTCCAGAGGCGGAAATC 1962  
Db |||||  
QY 1996 GCCAAGAGGGTTACGAAAATTTAGTTAGTCCCATACACATTTCTCCAGAGGCGGAAATC 2055  
QY 1963 CTTTTGAGGATATCCCAAGAAAACGGTTCCCGAAGTTTAAAGACATGATGTTAAC 2022  
Db |||||  
QY 2056 CTTTTGAGGATATCCCAAGAAAACGGTTCCCGAAGTTTAAAGACATGATGTTAAC 2115  
QY 2023 TCAACAAGGAGAGCCAGGAGAGTGCAAAATTCCTCCCTGGTAAATATTTTACCTCTCCA 2082  
Db |||||  
QY 2116 TCAACAAGGAGAGCCAGGAGAGTGCAAAATTCCTCCCTGGTAAATATTTTACCTCTCCA 2175  
QY 2083 AAAGACGCCAGTGTAGTCTCAATACCTTGGATTTGCAACTGGAACATGAGACATCACT 2142  
Db |||||  
QY 2176 AAAGACGCCAGTGTAGTCTCAATACCTTGGATTTGCAACTGGAACATGAGACATCACT 2235  
QY 2143 TTGAAAGGATACAAATTTGTCAGTCAAGCTTGCCTGAGATCATTTCTGATGAATCACAG 2202  
Db |||||  
QY 2236 TTGAAAGGATACAAATTTGTCAGTCAAGCTTGCCTGAGATCATTTCTGATGAATCACAG 2295  
QY 2203 CATCTAAAATAAAAATCAAGCTATAATAACAGATGAACAAATGACAGTTTGGTGGCT 2262  
Db |||||  
QY 2296 CATGCTAAAATAAAAATCAAGCTATAATAACAGATGAACAAATGACAGTTTGGTGGCT 2355  
QY 2263 CCACAGAAAAACAGGTTTCATAAAGCATCTTGCCAAAACAGCTTAGAGTGTCTGAAAGA 2322  
Db |||||  
QY 2356 CCACAGAAAAACAGGTTTCATAAAGCATCTTGCCAAAACAGCTTAGAGTGTCTGAAAGA 2415  
QY 2323 TTGACAGGTTGACTTTTCTGCAAGTGTGTAAGTGAATGGTCAATGACAGGCTCAG 2382  
Db |||||  
QY 2416 TTGACAGGTTGACTTTTCTGCAAGTGTGTAAGTGAATGGTCAATGACAGGCTCAG 2475  
QY 2383 AATCCACCCCTGGACCTGGAGACCAAGATTTAGAGTGAACAACTCACACCCAAAAA 2442  
Db |||||  
QY 2476 AATCCACCCCTGGACCTGGAGACCAAGATTTAGAGTGAACAACTCACACCCAAAAA 2535  
QY 2443 ACCATAGGGGAAATGTGACAAAAAGGCCCCATCTCTGATTTGCTTCCACTGGAAAGC 2502  
Db |||||  
QY 2536 ACCATAGGGGAAATGTGACAAAAAGGCCCCATCTCTGATTTGCTTCCACTGGAAAGC 2595  
QY 2503 CAGATGAC 2510

Db 2596 CAGATGAC 2603  
|||||||

## RESULT 5

AC010205

LOCUS

DEFINITION

AC010205

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

AC010205 162427 bp DNA linear PRI 13-AUG-2002  
Homo sapiens 12 BAC RP11-285E23 (Roswell Park Cancer Institute  
Human BAC Library) complete sequence.

AC010205

AC010205.5 GI:6468049

HTG.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 162427)

Muzny, D.M., Adams, C., Bailey, M., Barbara, J., Blankenburg, K.,

Bodota, B., Bock, J., Bowie, S., Brooks, A., Buhay, C., Bunac, C.,

Burkett, C., Burrows, J., Carter, M., Chacko, J., Chen, Z., Cox, C.,

David, R., Delgado, O., Deshazo, D., Ding, Y., Domah-Rashid, N.,

Dugan-Rocha, S., Durbin, K.J., Fernandez, C., Ferraguto, D.,

Forcum-Ransey, J., Frantz, P., Ganesh, R., Gorrell, J.H., Gorrell, L.L.,

Guevara, W., Harris, K., Hernandez, J., Hodgson, A., Hognes, M.,

Holloway, C., Hosak, H., Issar, A., Jackson, L.E., Jackson, L., Jia, Y.,

Jones, M., Kelly, S., Kneitz, S., Kondejewski, N., Kong, Y., Kovar, C.,

Lau, S., Leal, B., Lee, E., Li, Z., Lichtarge, O., Liu, J., Liu, W.,

Logan, O., Lu, J., Lucier, R., Marondel, I., Martin, R., Martinez, C.,

McLeod, M.P., Mei, G., Merscher, S., Miller, A., Montgomery, K.T.,

Morgan, M., Morris, S., Nash, S., Nelson, A., Nguyen, R., Nguyen, N.,

Nguyen, S., Oswal, G., Parish, B., Paxton, S., Payton, B., Perez, L.,

Pu, L.L., Quiles, M., Reiter, D., Rives, M., Samuel, S., Say, J.,

Scherer, S., Shah, E., Shen, H., Shim, C., Simon, M., Sparks, A.,

Stamps, A., Sugang, R., Tabor, P., Taylor, R., Vasquez, L., Vinson, R.,

Vo, O., Wabnah, M., Watlington, S., Weinstein, G., Weinstein, I.R.,

Williamson, A., Worley, K., Wren, J., Wrenford, G., Xiang, A.M.,

Yang, R., Yu, W., Zhou, X., Kucherlapati, R., Nelson, D. and Gibbs, R.

Direct Submission

Unpublished

2 (bases 1 to 162427)

Worley, K.C.

Direct Submission

Submitted (15-SEP-1999) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 162427)

Worley, K.C.

Direct Submission

Submitted (25-NOV-1999) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

4 (bases 1 to 162427)

Worley, K.C.

Direct Submission

Submitted (13-AUG-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Nov 25, 1999 this sequence version replaced gi:6087851.

INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email

gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the

entire insert of this clone. Overlapping regions of clones are only

sequenced and submitted once, so the sequence for the remainder of

the insert may be found in the record for the adjacent clones.

Overlapping clones are noted at the beginning and end of the

Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches

of a local database that includes entries from dbSTS, GDB, and

local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,

unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST  
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the  
EST and cDNA sequences. Genes demonstrate at least two exons  
flanked by consensus splice sites that maintained sequence  
continuity across the splice junctions. Sequences that are not  
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum  
standard of double strand coverage with a minimum of 2 clones and 2  
reads with no ambiguities or 2 chemistries with a minimum of 2  
clones and 3 reads with no ambiguities. If the sequence quality for  
a region does not meet this standard, it will be indicated in the  
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality  
standards - estimated error rate less than 1 per 10,000 bases.  
Reports of lowest quality individual bases and measures of base  
quality are listed below. Description of the metrics can be found  
at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

## QUALSTAT-REPORT-----

----- Summary Statistics -----  
Contig length: 144841  
Phrap values in estimate: 143745  
Average error rate (BCM-Phrap estimate): 0.000323023  
Fraction of Phrap values less than 40 : 0.052607  
Number of consensus changing edits: 22  
Number of N's in consensus : 0

----- Consensus changing edits -----  
Position Original+Context Edited+Context  
1530 atatatgaaa(n)tgagtagac atatatgaaa(t)tgagtagac  
3935 ctaaatgtga(n)tggtatttt ctaaatgtga(t)tggtatttt  
5613 tgtgttttg(n)ttctgttgtt tgtgttttg(t)ttctgttgtt  
25390 ggtttattg(n)aatatgaat ggtttattg(g)aatatgaat  
47091 ggtgaaaatt(n)ccctgtgtaa ggtgaaaatt(c)ccctgtgtaa  
65289 taatgcagca(t)tatccctct taatgcagca(a)tatccctct  
98027 gtcacagaag(n)ccacaatgta gtcacagaag(t)ccacaatgta  
105124 tctaatttaa(n)caagggaagt tctaatttaa(c)caagggaagt  
105232 aacagaatac(n)agacactggg aacagaatac(t)agacactggg  
105335 gctgcagct(n)agatnccactn gctgcagct(g)agatnccactn  
105340 atgctnagat(n)ccatnnctc atgctnagat(c)ccatnnctc  
105345 agatnccact(n)nnctcactgg agatnccact(g)nnctcactgg  
105346 gatnccactn(n)ctcactgga gatnccactn(c)ctcactgga  
105347 gatnccactn(n)ctcactgga gatnccactn(c)ctcactgga  
105371 ttccacttac(n)nnntnaaaat ttccacttac(a)nnntnaaaat  
105372 tccacttac(n)nnntnaaaat tccacttac(g)nnntnaaaat  
105373 tccacttacn(n)nnntnaaaat tccacttacn(c)nnntnaaaat  
105374 cacttacnn(n)tnaaataact cacttacnn(c)tnaaataact  
105376 ctatcnnnt(n)aaaataactga ctatcnnnt(c)aaaataactga  
105503 cctatcttt(n)atattagaat cctatcttt(c)atattagaat  
105569 ttttttttt(n)cttgagacag ttttttttt(t)cttgagacag  
112885 tctattaaga(n)atatacataa tctattaaga(a)atatacataa

## ----- Distribution of Quality &lt; 40 Bases -----

# bases	* * * * *
1000	* * * * *
900	* * * * *
800	* * * * *
700	* * * * *
600	* * * * *
500	* * * * *
400	* * * * *
300	* * * * *
200	* * * * *
100	* * * * *
0	* * * * *

-----									
	5	10	15	20	25	30	35	40	
	Phrap Value Range								
-----									
Version: 1.01 qxfo.									
Location/Qualifiers									
source	1. .162427								
	/organism="Homo sapiens"								
	/db_xref="taxon:9606"								
	/chromosome="12"								
repeat_region	351. .470								
	/rpt_family="AluJo/FLAM"								
repeat_region	complement(595. .683)								
	/rpt_family="MIR"								
repeat_region	complement(909. .1023)								
	/rpt_family="MER96B"								
repeat_region	complement(1024. .1324)								
	/rpt_family="Alusp"								
misc_feature	1174. .1554								
	/function="Low coverage"								
repeat_region	complement(1325. .1356)								
	/rpt_family="MER96B"								
repeat_region	1357. .1671								
	/rpt_family="ALUSx"								
repeat_region	complement(1672. .1723)								
	/rpt_family="MER96B"								
repeat_region	complement(1745. .1889)								
	/rpt_family="FLAM_C"								
repeat_region	1978. .2171								
	/rpt_family="MIR"								
repeat_region	2678. .2883								
	/rpt_family="L2"								
repeat_region	3235. .3367								
	/rpt_family="GC_rich"								
repeat_region	3470. .3493								
	/rpt_family="(CGG)n"								
repeat_region	complement(3735. .3864)								
	/rpt_family="FLAM_C"								
repeat_region	complement(3978. .4096)								
	/rpt_family="FLAM_C"								
repeat_region	4171. .4202								
	/rpt_family="(T)n"								
repeat_region	4275. .4405								
	/rpt_family="L2"								
repeat_region	complement(4706. .5004)								
	/rpt_family="Alusx"								
repeat_region	complement(5275. .5496)								
	/rpt_family="MIR"								
repeat_region	complement(5497. .5823)								
	/rpt_family="Alusg"								
repeat_region	5882. .6038								
	/rpt_family="MIR"								
repeat_region	complement(6225. .6524)								
	/rpt_family="Alusq"								
repeat_region	6672. .6806								
	/rpt_family="AluJo"								
repeat_region	6807. .7084								
	/rpt_family="AluJb"								
Query Match 30.7%; Score 1103.6; DB 9; Length 162427;									
Best Local Similarity 99.6%; Pred. No. 4.8e-278;									
Matches 1106; Conservative 0; Mismatches 4; Indels 0; Gaps 0;									
QY	1579	CAGATCATTTTCATGAATTTGATAAGTGATCCTTCTCCAAACACGACTCACTATATT	1638						
		I							

RESULT 6  
AC063950

AC003550/C	LOCUS	AC063950	174231 bp	DNA	linear	PRI 29-JAN-2002
	DEFINITION	Homo sapiens	12 BAC	RP11-511H9	(Roswell Park Cancer Institute Human	



BAC Library) complete sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE

AC063950  
AC063950.37 GI:18390143

HTG.  
Homo sapiens  
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
AUTHORS

1 (bases 1 to 174231)  
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-oshman, F.R., Allen, C.,  
Alsbury, S.L., Amaral, H.C., Are, J.R., Ayale, M., Banks, P.,  
Barbata, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D.,  
Bouck, J., Bowie, S., Brivava, M., Brown, E., Brown, M., Bryant, N.P.,  
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,  
Carroll, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
Chen, G., Chen, R., Chen, Z., Chiu, D., Chowdhry, I., Christopoulos, C.,  
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,  
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,  
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Emerling, S.,  
Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P.,  
Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N.,  
Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,  
Hamilton, K., Han, J., Harris, C., Harris, K., Hart, M., Havlak, P.,  
Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hoques, M.,  
Holloway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S.,  
Hume, J., Ioshikhes, I., Jackson, L.E., Jacobson, B., Jia, Y.,  
Johnson, R., Jolivet, S., Joudan, S., Karlsson, E., Kelly, S., Khan, U.,  
King, L., Korvah, J., Kovar, S., Kratovic, J., Kureshi, A., Landry, N.,  
Leal, B., Lee, E., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O.,  
Lieu, C., Liu, J., Liu, W., Loulseg, H., Lozano, R.J., Lu, X.,  
Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P.,  
Maroncel, I., Martin, R., Martindale, A., Martinez, E., Massey, E.,  
Mawhinney, E., McLeod, M.P., Meador, M., Mei, G., Merscher, S.,  
Metzker, M., Miller, A., Miner, G., Miner, Z., Mitchell, T.,  
Mohabbat, K., Montgomery, K.T., Morgan, M., Morris, S., Moser, M.,  
Neal, D., Nelson, D., Newton, J., Newton, N., Nguyen, N., Nguyen, N.,  
Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G.,  
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
Peters, L., Pickens, R., Primus, E., Pu, J.L., Quiles, M., Ren, Y.,  
Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savery, G.,  
Scherer, S., Scott, G., Shen, H., Shim, C., Shooshtari, N., Sisson, I.,  
Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H.,  
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R.,  
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K.,  
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Kucherlapati, R.,  
Weinstock, G. and Gibbs, R.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Direct Submission  
Unpublished  
2 (bases 1 to 174231)  
Worley, K.C.

Submitted (22-APR-2000) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

3 (bases 1 to 174231)  
Worley, K.C.

Submitted (29-JAN-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Jan 29, 2002 this sequence version replaced gi:17149338.  
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email  
[gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)

CLONE LENGTH: This sequence does not necessarily represent the  
entire insert of this clone. Overlapping regions of clones are only  
sequenced and submitted once, so the sequence for the remainder of  
the insert may be found in the record for the adjacent clones.

Overlapping clones are noted at the beginning and end of the  
Features listing.

#### ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches  
of a local database that includes entries from dbSTS, GDB, and  
local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,  
unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST  
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the  
EST and cDNA sequences. Genes demonstrate at least two exons  
flanked by consensus splice sites that maintained sequence  
continuity across the splice junctions. Sequences that are not  
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum  
standard of double strand coverage with a minimum of 2 clones and 2  
reads with no ambiguities or 2 chemistries with a minimum of 2  
clones and 3 reads with no ambiguities. If the sequence quality for  
a region does not meet this standard, it will be indicated in the  
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality  
standards - estimated error rate less than 1 per 10,000 bases.  
Reports of lowest quality individual bases and measures of base  
quality are listed below. Description of the metrics can be found  
at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

#### QUALSTAT-REPORT.

FEATURES	Source	Location/Qualifiers
STS	1..174231	/organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="12" /clones="RP11-511Hg" 132..402 /standard_name="133116"
repeat_region	626..927	/rpt_family="AluSx"
repeat_region	1405..1699	/rpt_family="AluSg"
repeat_region	2197..2330	/rpt_family="AluJb"
repeat_region	2334..2375	/rpt_family="A-rich"
repeat_region	3303..3346	/rpt_family="AT-rich"
repeat_region	3561..3860	/rpt_family="AluSx"
repeat_region	4930..4957	/rpt_family="AT-rich"
STS	6296..6422	/standard_name="97243"
STS	6356..6479	/standard_name="48565"
repeat_region	6607..6638	/rpt_family="AT-rich"
repeat_region	7063..7087	/rpt_family="AT-rich"
repeat_region	7594..7889	/rpt_family="AluSx"
STS	8562..8691	/standard_name="46931"
STS	8777..8893	/standard_name="3096"
repeat_region	complement(9277..9424)	/rpt_family="L2"
repeat_region	complement(9693..9967)	/rpt_family="L2"
repeat_region	complement(10118..10410)	/rpt_family="AluJb"





AUTHORS Fu, Y., Pan, H., D'Souza, A., Canfield, B. and Roe, B.A.  
TITLE Homo sapiens Chromosome 10 BAC Clone 14951 In I-CELL Disease Region  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 177364)  
AUTHORS Fu, Y., Pan, H., D'Souza, A., Canfield, B. and Roe, B.A.  
TITLE Direct Submission  
JOURNAL Submitted (11-AUG-1998) Department of Chemistry And Biochemistry,  
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
REFERENCE 3 (bases 1 to 177364)  
AUTHORS Fu, Y., Pan, H., D'Souza, A., Canfield, B. and Roe, B.A.  
TITLE Direct Submission  
JOURNAL Submitted (04-FEB-1999) Department of Chemistry And Biochemistry,  
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
REFERENCE 4 (bases 1 to 177364)  
AUTHORS Fu, Y., Pan, H., D'Souza, A., Canfield, B. and Roe, B.A.  
TITLE Direct Submission  
JOURNAL Submitted (06-FEB-1999) Department of Chemistry And Biochemistry,  
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
REFERENCE 5 (bases 1 to 177364)  
AUTHORS Fu, Y., Pan, H., D'Souza, A., Canfield, B. and Roe, B.A.  
TITLE Direct Submission  
JOURNAL Submitted (09-FEB-1999) Department of Chemistry And Biochemistry,  
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
REFERENCE 6 (bases 1 to 177364)  
AUTHORS Fu, Y., Pan, H., D'Souza, A., Canfield, B. and Roe, B.A.  
TITLE Direct Submission  
JOURNAL Submitted (27-FEB-1999) Department of Chemistry And Biochemistry,  
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
REFERENCE 7 (bases 1 to 177364)  
AUTHORS Fu, Y., Pan, H., D'Souza, A., Canfield, B. and Roe, B.A.  
TITLE Direct Submission  
JOURNAL Submitted (23-MAR-1999) Department of Chemistry And Biochemistry,  
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
REFERENCE 8 (bases 1 to 177364)  
AUTHORS Fu, Y., Pan, H., D'Souza, A., Canfield, B. and Roe, B.A.  
TITLE Direct Submission  
JOURNAL Submitted (24-MAY-2000) Department of Chemistry And Biochemistry,  
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
REFERENCE 9 (bases 1 to 177364)  
AUTHORS Fu, Y., Pan, H., D'Souza, A., Canfield, B. and Roe, B.A.  
TITLE Direct Submission  
JOURNAL Submitted (26-MAY-2000) Department of Chemistry And Biochemistry,  
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
REFERENCE 10 (bases 1 to 177364)  
AUTHORS Fu, Y., Pan, H., D'Souza, A., Canfield, B. and Roe, B.A.  
TITLE Direct Submission  
JOURNAL Submitted (27-MAY-2000) Department of Chemistry And Biochemistry,  
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
COMMENT On Feb 9, 1999 this sequence version replaced gi:4235169.  
FEATURES  
source  
Location/Qualifiers  
1. .177364  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/clone="14951"  
BASE COUNT 53772 a 37167 c 35988 g 50437 t  
ORIGIN  
Query Match 30.6%; Score 1103.2; DB 9; Length 177364;  
Best Local Similarity 99.7%; Pred. NO. 6.le-278;  
Matches 1105; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 1581 AGATCATTTTCATGAATTGTATAAAGTGATCCTTCTCCCAAAACAGAGTCTACTATATATAT 1640  
|||||

Db 51260 AGATCATTTTCATGAATTGTATAAAGTGATCCTTCTCCCAAAACAGAGTCTACTATATATAT 51201  
Qy 1641 TCCAAAAGGTGAATGCGCTTATTTTCAGCTTTGCAGAAAGTAGCCAAAAGAGAGTTGA 1700  
|||||  
Db 51200 TCCAAAAGGTGAATGCGCTTATTTTCAGCTTTGCAGAAAGTAGCCAAAAGAGAGTTGA 51141  
|||||  
Qy 1701 AGGTGCTTATAGTGACAAATCCAATAAATTCGACATGCTTCTATTGCCAACAAAGTGGAAC 1760  
|||||  
Db 51140 AGGTGCTTATAGTGACAAATCCAATAAATTCGACATGCTTCTATTGCCAACAAAGTGGAAC 51081  
|||||  
Qy 1761 CATCCACTCATATAGTACAGTGGAAATGCAATGCGCCACACATATATTTTATCTCAGTT 1820  
|||||  
Db 51080 CATCCACTCATATAGTACAGTGGAAATGCAATGCGCCACACATATATTTTATCTCAGTT 51021  
|||||  
Qy 1821 TCAAAATACAAACGATGAAGAGTTCAAAATGCAGATACAGTGGAGTGGACACAAAGGA 1880  
|||||  
Db 51020 TCAAAATACAAACGATGAAGAGTTCAAAATGCAGATACAGTGGAGTGGACACAAAGGA 50961  
|||||  
Qy 1881 GGGACCAAACTGAATCTACGGCCAGAGAGGTTAGGAAATTTAGTTAGTCCCATTAAC 1940  
|||||  
Db 50960 GGGACCAAACTGAATCTACGGCCAGAGAGGTTAGGAAATTTAGTTAGTCCCATTAAC 50901  
|||||  
Qy 1941 ACTTCTCCAGAGGGGAAATCTTTTGGAGATATTTCCCAAGAAAACCTTCCCGAA 2000  
|||||  
Db 50900 ACTTCTCCAGAGGGGAAATCTTTTGGAGATATTTCCCAAGAAAACCTTCCCGAA 50841  
|||||  
Qy 2001 GTTTAAGAGACATGATGTTAACTCAACAAGAGAGAGCCAGGAGAGTGAATAATCCCT 2060  
|||||  
Db 50840 GTTTAAGAGACATGATGTTAACTCAACAAGAGAGAGCCAGGAGAGTGAATAATCCCT 50781  
|||||  
Qy 2061 GGTAAATATTTCACTCTCTCCAAAAGAGCCAGTGTAGTCTCAATACCTTTGGATTGCA 2120  
|||||  
Db 50780 GGTAAATATTTCACTCTCTCCAAAAGAGCCAGTGTAGTCTCAATACCTTTGGATTGCA 50721  
|||||  
Qy 2121 ACTGGAACATGGAGACATCACTTTTGAAGAGATACAATTTGTCCAAGTACAGCTTGTGAG 2180  
|||||  
Db 50720 ACTGGAACATGGAGACATCACTTTTGAAGAGATACAATTTGTCCAAGTACAGCTTGTGAG 50661  
|||||  
Qy 2181 ATCATTTCTGATGAACTCACAGCATGCTAAATAAATAAATAAATAAATAAATAAATAAATAA 2240  
|||||  
Db 50660 ATCATTTCTGATGAACTCACAGCATGCTAAATAAATAAATAAATAAATAAATAAATAAATAA 50601  
|||||  
Qy 2241 AACAAATGACAGTTTGGTGGCTCCACAGGAAAAACAGGTTTCATAAAAGCATCTTGGCAAA 2300  
|||||  
Db 50600 AACAAATGACAGTTTGGTGGCTCCACAGGAAAAACAGGTTTCATAAAAGCATCTTGGCAAA 50541  
|||||  
Qy 2301 CAGCTTAGAGTGTCTGAAGATTCGAGAGTTGACTTTTCTGCAAGTGAATGAAGT 2360  
|||||  
Db 50540 CAGCTTAGAGTGTCTGAAGATTCGAGAGTTGACTTTTCTGCAAGTGAATGAAGT 50481  
|||||  
Qy 2361 GAATGGTATGACAGGTCAGAAATCCACCCCTGGACTTGGAGACACAGCAAGATTAG 2420  
|||||  
Db 50480 GAATGGTATGACAGGTCAGAAATCCACCCCTGGACTTGGAGACACAGCAAGATTAG 50421  
|||||  
Qy 2421 AGTGAAACTCACACCCCAAAACCATAGGCGGAATGTGACAAAGAAAAAGCCCCATC 2480  
|||||  
Db 50420 AGTGAAACTCACACCCCAAAACCATAGGCGGAATGTGACAAAGAAAAAGCCCCATC 50361  
|||||  
Qy 2481 TCTGATTTTCCACTGGAAAGCCAGATGACAAAAGAAAAGAAAATTCACAGGGAAGAAA 2540  
|||||  
Db 50360 TCTGATTTTCCACTGGAAAGCCAGATGACAAAAGAAAAGAAAATTCACAGGGAAGAAA 50301  
|||||  
Qy 2541 AGAGAACAGTAGAATGGAGGAAAATGCTGAAAATCACATAGCGCTTACTGAAGTCTTACT 2600  
|||||  
Db 50300 AGAGAACAGTAGAATGGAGGAAAATGCTGAAAATCACATAGCGCTTACTGAAGTCTTACT 50241  
|||||  
Qy 2601 TGGAAAGAGCTGCAGCATTACACAGATAGTTACTTGGGCTTTTGGCCATGGGAGAAAA 2660  
|||||  
Db 50240 TGGAAAGAGCTGCAGCATTACACAGATAGTTACTTGGGCTTTTGGCCATGGGAGAAAA 50181  
|||||  
Qy 2661 AAAGTATTTCTTAGATCTTCTCGACGAA 2688  
|||||  
Db 50180 AAAGTATTTCCAAGATCTTCTCGACGTA 50153  
|||||

```

RESULT 8
MUSBDLZ
LOCUS      1846 bp      mRNA      linear      ROD 02-MAY-1995
DEFINITION Mus Musculus basic domain/leucine zipper transcription factor mRNA,
3' end of cds.
ACCESSION L36434.1 GI:625041
VERSION    L36434
KEYWORDS   basic domain/leucine zipper transcription factor.
SOURCE     Mus musculus cDNA to mRNA.
ORGANISM   Mus musculus
REFERENCE  1 (bases 1 to 1846)
AUTHORS    Cordes, S.P. and Barsh, G.S.
TITLE      The mouse segmentation gene kr encodes a novel basic domain-leucine
zipper transcription factor
JOURNAL    Cell 79 (6), 1025-1034 (1994)
MEDLINE    95094266
PUBMED     79094266
FEATURES   Location/Qualifiers
            1..1846
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            1..1846
            <1..1155
            /codon_start=1
            /product="basic domain/leucine zipper transcription
factor"
            /protein_id="AA65688.1"
            /db_xref="GI:625042"
            TGRFEVEVKIPRVNLSLLPEAOLVRLNLDLQLEGGITLKGYNLSKSLRFLGNS
            LDWKIKQARTDETGNLEVPQENPSRRPHGRSEWTPAPATVTVKGRDHAL
            NPPVLNENRLEAQTGLTVPSKRNLSPLVPESHLPKEESDRAGNAPVPKVELPV
            GRCSRRIQAFPCPKKKYFDLLDAEESLQTLAYFTDSKTKRGLKDFADSLRYV
            NKIUNSPGTSRVPKPMHMDIRVMDQDMFPEFDKTSFKVHSEDMQFAFS
            YFYILMSAVOPLNISQVFEHVDTDQSVLSDREIRTLRIHLPLTCRT"
            1156..1846
3'UTR      polyA_site 572 a 451 c 415 g 401 t 7 others
BASE COUNT
ORIGIN
Query Match      23.6%; Score 849.8; DB 10; Length 1846;
Best Local Similarity 75.1%; Pred. No. 1.6e-211;
Matches 1257; Conservative 0; Mismatches 341; Indels 75; Gaps 13;
QY 1900 ACGGCCAGAGAGGGTTACGAAATTTAGTTAGTCCCATACACTTCTTCCAGAGCGCGAA 1959
DB 7 ACACCCAGAGGGCCTATGAAGTTTGGTTAGCCCGAGTGACACCTCTTCTCCAGGCTGAC 66
QY 1960 ATCCCTTTTGGAGGATATCCCAAGAAAGCGCTTCCCGAAGTTTAAAGAGACATGATGTT 2019
DB 67 GTCCCTTTTGAAGATGTCCTCCCAAGAGAAAGCGCTTCCCAAGATCAGGAGACATGATGA 126
QY 2020 AACTCAACAGAGAGCGCCAGAGAGAGTGAATAATTCCTCGGPAATAATTTTCACTCCTT 2079
DB 127 AATGCACAGGAGGATTTCAAGAGAGGAGTGAATAATTCCTCGGPAATAATTTTCACTCCTT 186
QY 2080 CCAAAAGACGCCAGTTCAGTCTCAATACCTTGGATTTGCAACTGGAACTGGAGACATC 2139
DB 187 CCCAAGAGGCCAGGCTGAGGCTGAGCAACTTGGATTTGCAACTGGAGAGATC 246
QY 2140 ACTTTGAAGGATACAAATTTGTCCAAAGTCAGCCTTGTGTGAGATCATTTCTTGATGAACATCA 2199
DB 247 ACTCTGAAAGGATATACTTTGCCAAGTCAGCCCTGCTAAGCTTTTCTTGGGGAATTCAC 306
QY 2200 CAGCATGCTAAATAAATAACAGCTATATAACAGATGAACAAATACAGCTTTGGT 2259
DB 307 CTAGATATAAATAAATAACCTCAAGCT---AGGACCGATGAACAAAGGCAACCTGGAG 363
QY 2260 GCTCCACAGGAAACAGGTTTCAATAAAGCATCTTGCACCAACAGCTTAGGAGTCTCTGAA 2319

```

```

Db 364 GTCCACAGGAAACCCCTTCTCAGAGC-----TCCACATGGCTTGTGTGACAC 417
QY 2320 AGATTGCAGAGTTGACTTTTCTCGCAGTGAGTGTAAAGTGAATGTCTATGACAGGGT 2379
Db 418 AGATCAGAGAGATGGACTGCCCGCAGCAGACAGTGCAGCGTGAAGCGCGTGACACGCT 477
QY 2380 CAGAAATCCACCCCTGGAGTTGGAGACCACAGCAAGATTTAGAGTGGNAACTCACACCCA 2439
Db 478 TTGAATCCACCCCGGTGTTGGAGACCAATGCAAGATTGG-----CCGAG 522
QY 2440 AAAACCATAGCGGAAATGTGACAAAAGAAAGCCGCCCTCTCTGATTTGTTCCACTGGAA 2499
Db 523 CCTACACTAGCGGTGACTGTGTCCAAAGAGAACTTTCACCGCTGATGCTGTCCCCAGAA 582
QY 2500 AGCCAGATACAAAAGAAAAGAAAATACAGAGGAAAGAAAAGAACAGACAGTAGAATGGAG 2559
Db 583 AGCCACTTGCCTGCAAAAGAGAGAGAGTGCAGGGCAGAA----- 621
QY 2560 GAAATGCTGAAAATCACATAGGCGTTACTGAAGTGTACTTGGAGAAAGCTGCAGCAT 2619
Db 622 -----GGCAATGCTGTACCTGTAAAGGAGTTAGTGCCTGGCAGCG----- 663
QY 2620 TACACAGATAGTTACTTGGCGTTTTTGGCCATGGGAGAAAAGATATTTCCTAGATCTT 2679
Db 664 TGCAGCAGAAATTTATCCAGGCTTTTGGCCCTGGGAAAAGAAAAGATTTTCCAAAGACCTT 723
QY 2680 CTGCAGCAAGAGATCATTTGAAGACACATTTGGCCCTACTCTACTGATAGCAAGATAGA 2739
Db 724 CTTGTCGGGAGAGTCAATTTGAAGACCCAGTGTGGCGTACTTTTACAGACAGCAACATACC 783
QY 2740 GCCAGATACAAAGAGATACATTTGCAGATTTCCCTCAGATATGTAAATAAATTTCTAAAT 2799
Db 784 GGGAGGCACTAAAGATACATTTGCAGACTCCTCCGATACGTCAATAAATTTCTCAAC 843
QY 2800 AGCAAGTTTGGATTACATCGCGGAAAGTCCCTGCTCAGATGCTCAGATGATTTGACCG 2859
Db 844 AGCAAGTTTGGATTACATCGCGGAAAGTCCCTGCTCAGATGCTCAGATGATTTGACAG 903
QY 2860 ATTGTTATGCAACACTCGAAGATATGTTCCCTGGAAGATTTGCAAGACGTCATTTCAC 2919
Db 904 ATCGTTATGCAAGACTCCAGATATGTTCCCTGGAAGATTTGCAAGACTTCATTTCAC 963
QY 2920 AAGTGCAGCATTTGAGGATATGAGTTTGGCTTCTCTTATTTTATTTATTCATGAT 2979
Db 964 AAGTGCAGCATTTGAGGATATGAGTTTGGCTTCTCTTATTTTATTTATTCATGAT 1023
QY 2980 GCAGTGCAGGCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGT 3039
Db 1024 GCAGTGCAGGCACTGAATATATTTCCCAAGTCTTTTCATGAAGTAGACACAGACCAATCTGT 1083
QY 3040 GTCTTGTCTGACAGAGAAATCCGAACACTGGCTACCAGAAATTCAGAACTGCGGTAGT 3099
Db 1084 GTCTTGTCTGATAGGAAATCCGAACACTGGTCCACAGAAATTCAGACCTACTCTTTAA-C 1142
QY 3100 TTGAGGATTTGACAGGCTGGAACACATGCTAATAAATTCG-TCAAAAATCTTCTCTGC 3158
Db 1143 TTGAGGATTTGACAGGCTTGGAAACACATGTTAATAAATTCGNNCAAAAATTCNCCCGC 1202
QY 3159 TGATATCAGCAGCTAAATAATTTCCACCACTCAGGAATCCTACTATGATCCCAAGCT 3218
Db 1203 TAATAT-----CAACTCAACAACTCCACCGACTCAGGAAGCATACTAGACCCCAAGCT 1258
QY 3219 GCCACCGGT-CACATAAAGCTAGTAACAACTGTAAACACTGTAAACACTGTACAAAATCCACA 3277
Db 1259 GCCTCGGTCCTACTAGAGTCTTGTCCACCACTGT--ACCAGTAAGTACAGATGCCACA 1316
QY 3278 AAGCATATAGGACAAAAAATAATATAGGTTTGAATCATGGGAGAAAGAAATTCGCTT 3337
Db 1317 AAGCCCTATAA-GACAAGAAATAATACAGGTTTGAACACTCATGGGAGAGGAAGAAAGCTT 1375
QY 3338 TTAATATGATTCGTACCAACGTTTCTCATGTGGTGGCCAGTTGGATGACATAGAAGAAA 3397

```

Db 1376 CAAGATGATACAAACCAAGTGTCTTCTCATGTGCTGGCCAGTTGGATGACATCAGAAAAA 1435

QY 3398 ACCTAGGAAGTTTGTTCCTGAATGACACAACTTACACAACTATCAAGATG-CTCAG 3456

Db 1436 ACCCCAGGAAGTTGCTGTGAATGACAAAAATTGACCAACCAAGATGCCCCGG 1495

QY 3457 ACAGTGAAGCTGTCTTCAGGGACTTCTATGAATCCATGTTCGCCATACCTTCCCAATTT 3516

Db 1496 ACAGTGAAGCTGTCTTCAGGGACTTCTATGAGTCCATGTTCGCCATACCTTCCCAAGTT 1555

QY 3517 GAAGTCCAGAGAGATGATCAAGACCGTTTCCCTTCATATGATGATGAGTGCAGGA 3569

Db 1556 GNGCTGCCAAGAGTATCGAAGCCGCTTTC-TGACATGATGATGAGTCCAAAGA 1607

RESULT 9

LSM802690

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 2428)

Bloeker, H., Boecher, M., Brandt, P., Mewes, H.W., Weil, B. and Wiemann, S.

Direct Submission

Submitted (15-JUN-2000) MIPS, Am Klopferspitz 18a, D-82152 Martinsried, GERMANY

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by GfR (National Research Centre for Biotechnology Ltd., Braunschweig/Germany) within the cDNA sequencing consortium of the German Genome Project.

This clone (DKFZp762B226) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at <http://www.mips.biochem.mpg.de/proj/cDNA/>.

FEATURES

source

1..2428

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/map="496.6 cr from top of Chr12 linkage group"

/clone="DKFZp762B226"

/tissue\_type="melanoma (MeWo cell line)"

/clone\_lib="762 (synonym: hmel2). Vector pSport1; host DH10B; sites NotI + SalI"

/dev\_stage="adult"

1..747

/gene="DKFZp762B226"

<1..747

/gene="DKFZp762B226"

/note="similarity to KIAA1208"

/codon\_start=1

/product="hypothetical protein"

/protein\_id="CAB94874.1"

/db\_xref="GI:8655650"

/translations="PTRPVDFVDVTDQSGVLSDPRETRTLATRIHELPLSLQDLTGLEH MLNCKMLPADITQLNIPPTQESYDNLPTVTSKLVNCKPVTDKHAYKKNK YREIFGEESIEAFKMTNPNVSHVQGLDDIRKNPRFVCLINDINHHKDAQTVRAVL RPYSEMFPPSPFELPREYRFLHHELOEMRAYRDLKWLKTHCVLATLIMFTIFS FFAEQIALRKRPFRRIHKEASPNRIY"

polyA\_site 725 a 483 c 460 g 760 t

BASE COUNT

ORIGIN

Query Match 16.5%; Score 592.8; DB 9; Length 2428;

Best Local Similarity 99.7%; Pred. No. 4.1e-144;

Matches 594; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

RESULT 10

AC121402/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 141215)

Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C., Alsbrooks, S.L., Amaral, H.C., Are, J.R., Ayelle, M., Banks, T., Barbra, J., Benton, J., Binage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, F., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,

AC121402

Rattus norvegicus clone CH230-324F15, \*\*\* SEQUENCING IN PROGRESS

\*\*\*, 42 unordered pieces.

AC121402

AC121402.2 GI:21903099

HTGS\_PHASE1.

Rattus norvegicus

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 141215)

Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C., Alsbrooks, S.L., Amaral, H.C., Are, J.R., Ayelle, M., Banks, T., Barbra, J., Benton, J., Binage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, F., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,

Homai, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L. E.,  
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,  
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,  
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L.,  
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulsegad, H.,  
 Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,  
 Massey, E., Mawhney, E., McLeod, M. P., Meador, M., Mei, G., Metzker, M.,  
 Miner, G., Minter, G., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okuonu, G.,  
 Oragunye, N., Oviado, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
 Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y.,  
 Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savery, G.,  
 Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,  
 Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H.,  
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
 Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,  
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
 Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,  
 Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstein, G. and Gibbs, R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 141215)  
 Worley, K. C.  
 Direct Submission  
 Submitted (18-MAY-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 141215)  
 Worley, K. C.  
 Direct Submission  
 Submitted (23-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jul 18, 2002 this sequence version replaced gi:20976340.  
 -----  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 -----  
 Center project name: CXAO  
 Center clone name: CH230-324F15  
 -----  
 Summary Statistics  
 -----  
 Sequencing vector: Plasmid;  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 108630 bases at least Q40  
 Consensus quality: 111787 bases at least Q30  
 Consensus quality: 114494 bases at least Q20  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 42 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \*  
 \* 1223: contig of 1223 bp in length  
 \* 1224: 1323: gap of unknown length  
 \* 1324: 2695: contig of 1372 bp in length  
 \* 2696: 2795: gap of unknown length  
 \* 2796: 3942: contig of 1147 bp in length  
 \* 3943: 4042: gap of unknown length  
 \* 4043: 5211: contig of 1169 bp in length  
 \* 5212: 5311: gap of unknown length  
 \* 5312: 6586: contig of 1275 bp in length  
 \*  
 \* 6587: gap of unknown length  
 \* 6588: 7910: contig of 1224 bp in length  
 \* 7910: 8010: gap of unknown length  
 \* 8011: 9051: contig of 1041 bp in length  
 \* 9052: 9151: gap of unknown length  
 \* 9152: 10546: contig of 1395 bp in length  
 \* 10547: 10646: gap of unknown length  
 \* 10647: 11871: contig of 1225 bp in length  
 \* 11872: 11971: gap of unknown length  
 \* 11972: 13687: contig of 1716 bp in length  
 \* 13688: 13787: gap of unknown length  
 \* 13788: 15781: contig of 1994 bp in length  
 \* 15782: 15881: gap of unknown length  
 \* 15882: 17949: contig of 2068 bp in length  
 \* 17950: 18049: gap of unknown length  
 \* 18050: 19781: contig of 1732 bp in length  
 \* 19782: 19881: gap of unknown length  
 \* 19882: 21439: contig of 1558 bp in length  
 \* 21440: 21539: gap of unknown length  
 \* 21540: 22815: contig of 1276 bp in length  
 \* 22816: 22915: gap of unknown length  
 \* 22916: 24651: contig of 1736 bp in length  
 \* 24652: 24751: gap of unknown length  
 \* 24752: 27211: contig of 2460 bp in length  
 \* 27212: 27311: gap of unknown length  
 \* 27312: 29806: contig of 2495 bp in length  
 \* 29807: 29906: gap of unknown length  
 \* 29907: 32159: contig of 2263 bp in length  
 \* 32170: 32289: gap of unknown length  
 \* 32290: 34289: contig of 2020 bp in length  
 \* 34290: 34389: gap of unknown length  
 \* 34390: 37438: contig of 3049 bp in length  
 \* 37439: 37538: gap of unknown length  
 \* 37539: 40323: contig of 2785 bp in length  
 \* 40324: 40423: gap of unknown length  
 \* 40424: 43380: contig of 2957 bp in length  
 \* 43381: 43480: gap of unknown length  
 \* 43481: 46912: contig of 3432 bp in length  
 \* 46913: 47012: gap of unknown length  
 \* 47013: 50378: contig of 3366 bp in length  
 \* 50379: 50478: gap of unknown length  
 \* 50479: 52969: contig of 2491 bp in length  
 \* 52970: 53070: gap of unknown length  
 \* 53070: 56475: contig of 3406 bp in length  
 \* 56476: 56575: gap of unknown length  
 \* 56576: 59055: contig of 2480 bp in length  
 \* 59056: 59155: gap of unknown length  
 \* 59156: 63855: contig of 4700 bp in length  
 \* 63856: 63955: gap of unknown length  
 \* 63956: 68690: contig of 4735 bp in length  
 \* 68691: 68790: gap of unknown length  
 \* 68791: 72439: contig of 3649 bp in length  
 \* 72440: 72539: gap of unknown length  
 \* 72540: 74492: contig of 1953 bp in length  
 \* 74493: 74592: gap of unknown length  
 \* 74593: 79801: contig of 5209 bp in length  
 \* 79802: 79901: gap of unknown length  
 \* 79902: 83391: contig of 3490 bp in length  
 \* 83392: 83491: gap of unknown length  
 \* 83492: 86836: contig of 3345 bp in length  
 \* 86837: 86936: gap of unknown length  
 \* 86937: 91723: contig of 4787 bp in length  
 \* 91724: 91823: gap of unknown length  
 \* 91824: 98085: contig of 6262 bp in length  
 \* 98086: 98185: gap of unknown length  
 \* 98186: 105770: contig of 7585 bp in length  
 \* 105771: 105870: gap of unknown length  
 \* 105871: 111548: contig of 5678 bp in length  
 \* 111549: 111648: gap of unknown length  
 \* 111649: 120177: contig of 8329 bp in length  
 \* 120178: 120277: gap of unknown length  
 \* 120278: 129704: contig of 9427 bp in length  
 \* 129705: 129804: gap of unknown length

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

COMMENT

\* 129805 141215: contig of 11411 bp in length.

FEATURES  
Source 1..141215  
Location/Qualifiers  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
/clone="CH230-324F15"

BASE COUNT 38123 a 29118 c 28492 g 38666 t 6816 others  
ORIGIN

Query Match 13.98; Score 499; DB 2; Length 141215;  
Best Local Similarity 69.6%; Pred. No. 1.8e-119;  
Matches 777; Conservative 0; Mismatches 255; Indels 85; Gaps 4;

```
QY 1572 CTGTGGGCAAGATCATTTTCATGAATGTATATAAGTGCCTCTCCCTCCCAACAGACTCA 1631
DB 62137 CTTTCCTCTAGATCATTTTCATGAATGTATATAAGTGCCTCTCCCTCCCAACAGACTCA 62078

QY 1632 CTATATATTCCAAAGGTGAATGCCCTTATTTTCAGCTTTTCGAGAAGTAGCCAAAG 1691
DB 62077 CTATGTCTGCCCAAGGTGAATACCTGTCTTATTTTCAGCTTTTCGAAACATAGCCAAAG 62018

QY 1692 AGGAGTTGAAGGTGCCATATGACATCCATTAATTCGACATGCTTCTATTGCCAAACAA 1751
DB 62017 AGGGGTTGAGGGGACCTACAGTGCACACCCCATCGCTCGTCACGCGTCTATCGCAACAA 61958

QY 1752 GTGGAACACCATCCACCTCATATGTCACAGTGGAAATGAATGCCACCAACATACATTTAA 1811
DB 61957 GTGGAACACCATACCTCATATGTCACAGTGGGATGAACCAACCAAGTCTATTTTAA 61898

QY 1812 TCTCAGCTTTCAAATCAACAGTGAAGAGTTCAAAATGCAGATAACAGTGGAGTGA 1871
DB 61897 CTTCACTCTTCAAACCTCAACAGTGAAGAGTTCAAAATGCAGATAACAGTGGAGTGA 61838

QY 1872 CACAAGGAGGACCAACATGAATTTCTACGGCCCAAGGTTTACGAAATTTAGTTAG 1931
DB 61837 CACGAGGAGGTGCTTAACCGAATCTCAACACCCAGAAGACCTTATGACAAATTTGTTAG 61778

QY 1932 TCCATAACACTCTTCCAGAGCGGGAATCTTTTTCAGATATTCCTCCAAAGAAACG 1991
DB 61777 CCCAGTGACACTCTTCTCAGGTGAGTCCCTTTTGAAGATGTTCCCAAGAGAAAG 61718

QY 1992 CTTCCCGAAGTTTAAAGACATGATGTTAACTCAACAGGAGAGCCAGGAAGAGTGA 2051
DB 61717 CTTCCCGAAGTTTAAAGACATGATGTTAAAGTCAACAGGAGAGTGA 61658

QY 2052 AATTCCCTGTGAATATTTTCACTCTTCCAAAGAGCCCGTGTAGTCTCAATACCTT 2111
DB 61657 AATCCCGGTGAATATTTTCCCTCTTCCAAAGAGCCCGTGTAGTGTAGTGTAGCAACCT 61598

QY 2112 GGATTTGCACTGGAACATGAGACATCACTTTTGAAGAGTATCAATTTGTCCAAAGTCAGC 2171
DB 61597 GGATTTGCACTGGAACATGAGACATCACTTTGAAAGAGTATCAATTTGTCCAAAGTCAGC 61538

QY 2172 CTTGCTGAGATCATTTCTGATGACTCACAGATGCTTAAATAAATAAATCAAGCTATAAT 2231
DB 61537 CTTGCTGAGATCATTTCTGATGACTCACAGATGCTTAAATAAATAAATCAAGCTATAAT 61482

QY 2232 AACAGATCAACAAATGACAGTTTGTGCTTCCACAGAGAAACAGCTTTCATTAAGCAT 2291
DB 61481 GACAGATGACAAAGAGCAACCTTGGAGTCCACAGAGAGAGCCCTCTCACCACAGC--- 61425

QY 2292 CTTGCCAACAATAGGAGTGTCTGAAAGATTCAGAGGTGTACTTTTCTGAGTGAG 2351
DB 61424 ---TCCAGGTGCTTTGCTGCCGACCGACAGACAGAGGTGGACGACCGGAGAC 61368

QY 2352 TGTAAGTGAATGGTCATGACAGGTGCAAGATCCACCCCTGGACTTGGAGACACAGC 2411
DB 61367 AGCGAGTGTGAACGGTGGTACACCGGTTTGAATCCACCCCG----- 61325

QY 2412 AAGATTTAGTGGAACTCACACCCCAAAACCATAGGCGAAATGTGACAAAAGAAA 2471
DB 61324 -----ACCAACCTGTGCCCAAGAGAA 61302
```

```
QY 2472 GCCCCCATCTCTGATTGTTCCACTGGAAGCCAGATGACAAAGAAAGAAATCACAGG 2531
DB 61301 CTTCTCACCACATGATCATTTCCCCAGAGCCGATGCAAGAGGAGG----- 61253

QY 2532 GAAAGAAAAAGACACAGTAGAATGGAGAAAAATGCTGAAAAATCACATAGGCGTTACTGA 2591
DB 61252 -----GACAGAGACAGACAGAGGCGCATGCTGATGCCAGGTACTGTAAACGA 61203

QY 2592 AGTGTACTTGGAGAAAGCTGCAGCATTACACAGATATTTACTTGGGCTTTTGGCCATG 2651
DB 61202 GCTGTGCTTGGCAGGAGGTACAGCAGCAGCATACCCAGTACCAGGCTTTTGGCCCTG 61143

QY 2652 GGAGAAAAAAGTATTTCTAGATCTTCTCGACGAA 2688
DB 61142 GGAGAAAAAAGTATTTCTCAAGACCTTCTTGATGTA 61106

RESULT 11
AC125486 281892 bp DNA linear HTG 27-JUN-2002
LOCUS Mus musculus chromosome UNK clone RP23-21G7, WORKING DRAFT
DEFINITION SEQUENCE, 15 unordered pieces.
ACCESSION AC125486
VERSION AC125486.1 GI:21617767
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 281892)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 281892)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (27-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: M_BA0021G07
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 274975 bases at least Q40
Consensus quality: 276824 bases at least Q30
Consensus quality: 278008 bases at least Q20
Insert size: 230000; agarose-fp
Insert size: 280492; sum-of-contigs
Quality coverage: 11.26 in Q20 bases; agarose-fp
Quality coverage: 7.91 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1622: contig of 1622 bp in length
* 1623 1722: gap of unknown length
* 1723 4017: contig of 2295 bp in length
* 4018 4117: gap of unknown length
```





/tissue\_type="cdna-collection"  
/clone\_lib="686 (synonym: hicc3). Vector pSport1\_Sfi; host  
DH10B; sites SfiIA + SfiIB"  
/dev\_stage="adult"  
polyA\_signal 2246..2251  
polyA\_site 2256  
BASE COUNT 670 a 462 c 468 g 732 t  
ORIGIN

Query Match 11.2%; Score 403.8; DB 9; Length 2332;  
Best Local Similarity 99.5%; Pred. No. 1.5e-94;  
Matches 405; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 337 GAGAAGCAGTTAGAGTGTTCCTAACACACTGCATTAAAGTGCCAATGCTTGTCTCTGGAC 396  
Db 1 GAGAAGCAGTTAGAGTGTTCCTAACACACTGCATTAAAGTGCCAATGCTTGTCTCTGGAC 60  
QY 397 CGAGCCCTGCCGCCACATCACCTGAAGGAGCTGCCATCTCTTTATCCTCTCTTTTCAT 456  
Db 61 CGAGCCCTGCCGCCACATCACCTGAAGGAGCTGCCATCTCTTTATCCTCTCTTTTCAT 120  
QY 457 TCTGCCAGTGCATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGCTCAGTT 516  
Db 121 TCTGCCAGTGCATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGCTCAGTT 180  
QY 517 GTTGTCTTTCACAGTACTAAGAGTGTGAAGATGCCCACTCTGGACTGCTTAAAGGAAAT 576  
Db 181 GTTGTCTTTCACAGTACTAAGAGTGTGAAGATGCCCACTCTGGACTGCTTAAAGGAAAT 240  
QY 577 AGCAGACAGACATGAGGGGCTACTTCACACAGATAAAGAGTCCCTGGATTAGTG 636  
Db 241 AGCAGACAGACATGAGGGGCTACTTCACACAGATAAAGAGTCCCTGGATTAGTG 300  
QY 637 CTAATGCAAGATTGGCTTTCCTGAGTGGATTTCACACACATTCGAAGAAACAATCAA 696  
Db 301 CTAATGCAAGATTGGCTTTCCTGAGTGGATTTCACACACATTCGAAGAAACAATCAA 360  
QY 697 CTAACAAACAAATGGCAGAAATCTTCTCTTAAAGTCAAACTGTT 743  
Db 361 CTAACAAACAAATGGCAGAAATCTTCTCTTAAAGTCAAACTGTT 407

RESULT 13  
AK001821  
LOCUS Homo sapiens cDNA FLJ10959 fis, clone PLACE1000562. PRI 01-AUG-2002  
DEFINITION AK001821  
ACCESSION AK001821  
VERSION AK001821.1 GI:7023328  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens placenta cDNA to mRNA, clone\_lib: PLACE1  
clone: PLACE1000562.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1  
Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,  
Nishikawa, T., Nagai, K., Sugano, S., Takahashi-Fujii, A., Hara, H.,  
Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hata, R., Takeuchi, K.,  
Arita, M., Nabekura, T., Ishii, S., Kawai, Y., Saito, K., Yamamoto, J.,  
Wakamatsu, A., Nakamura, Y., Nagahari, K., Masuho, Y. and Oshima, A.  
NEDO human cDNA sequencing project

Unpublished  
2 (bases 1 to 2076)  
Isogai, T. and Otsuki, T.  
Direct Submission  
Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute,  
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)  
NEDO human cDNA sequencing project supported by Ministry of  
International Trade and Industry of Japan; cDNA full insert  
sequencing; Research Association for Biotechnology; cDNA library  
construction, 5'- & 3'-end one pass sequencing and clone selection;  
Helix Research Institute (supported by Japan Key Technology Center

etc.) and Department of Virology, Institute of Medical Science,  
University of Tokyo.  
FEATURES  
source Location/Qualifiers  
1..2076  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="PLACE1000562"  
/tissue\_type="placenta"  
/clone\_lib="PLACE1"  
/note="Cloning vector: pME18SFL3"  
11..409  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="BAA91926.1"  
/db\_xref="GI:7023329"  
/translation="MIRTNVSHVVGQLDDIRKNPRKRVCLNDINIDHNHDKAQIVKAVL  
RDFYSEMFPIPSKIFPREYRNLHMHLEQWRAWDKLFKFWHCVLATLIMFTIFS  
FFAQQLIALKRKIFPRRRIRHKEASPNRIRV"  
BASE COUNT 588 a 409 c 400 g 679 t  
ORIGIN

Query Match 7.4%; Score 266; DB 9; Length 2076;  
Best Local Similarity 100.0%; Pred. No. 2.1e-58;  
Matches 266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 3333 CGCTTTTAAATGATTTCGTACACAGTTTCTCATGTGCTGCGCCAGTTGGATGACATAAG 3392  
Db 1 CGCTTTTAAATGATTTCGTACACAGTTTCTCATGTGCTGCGCCAGTTGGATGACATAAG 60  
QY 3393 AAAAACCCCTAGGAAGTTTGTTCCTGAATGACACAACTTGACCACAACTATAAGATGC 3452  
Db 61 AAAAACCCCTAGGAAGTTTGTTCCTGAATGACACAACTTGACCACAACTATAAGATGC 120  
QY 3453 TCAGACAGTGAAGCTGTTCTCAGGAGCTTCTATGAATFCCATGTTCCCATACCTCCCA 3512  
Db 121 TCAGACAGTGAAGCTGTTCTCAGGAGCTTCTATGAATFCCATGTTCCCATACCTCCCA 180  
QY 3513 ATTTGAACCTGCAAGAGAGTATCGAAACCGTTTCTTCATATGATGATGAGCTGACAGGAATG 3572  
Db 181 ATTTGAACCTGCAAGAGAGTATCGAAACCGTTTCTTCATATGATGATGAGCTGACAGGAATG 240  
QY 3573 GAGGCTTATCGACACAAATTTGAAGT 3598  
Db 241 GAGGCTTATCGACACAAATTTGAAGT 266

RESULT 14  
AL135758  
LOCUS Mouse DNA sequence from clone RP21-117J5 on chromosome 5, complete  
DEFINITION AL135758 142667 bp DNA linear ROD 29-JUN-2002  
sequence.  
ACCESSION AL135758  
VERSION AL135758.32 GI:9799912  
KEYWORDS HTG.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 142667)  
Blakey, S.  
Direct Submission

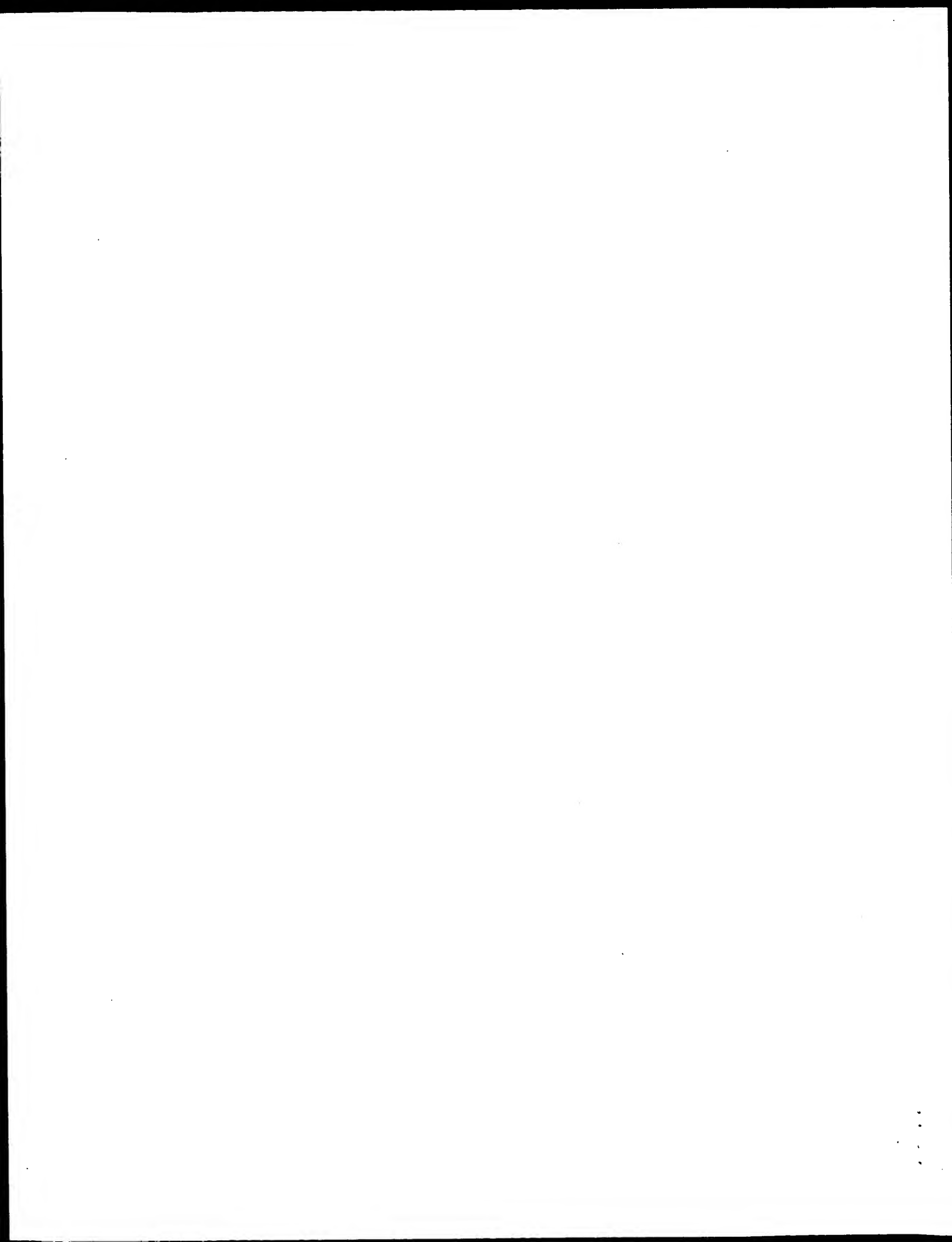
Submitted (29-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk  
On Aug 13, 2000 this sequence version replaced gi:8978010.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:





QY 3061 CGAACACTGGCTACCGAATTTCAGAACTCCCGTTAAGTTTCAGGATTTCACAGGCTCG 3120  
Db 10160 CGAACACTGGCCACGAGAAATTCACGACCTACCTTTAAGCTTGCAAGTAAAGTATTGTTTT 10101  
QY 3121 GAACACATGCTAATAAATTG 3140  
Db 10100 AACCTTACGCTTGTGACGTG 10081

Search completed: April 13, 2003, 01:14:45  
Job time : 8201 secs



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2003, 03:20:41 ; Search time 29.9002 Seconds  
(without alignments)  
1179.859 Million cell updates/sec

Title: US-10-023-888-2  
Perfect score: 6340  
Sequence: 1 METDTLLWLLVLLVPGSTG.....NRELMHQLQEWAYRDKLK 1199

Scoring table: BLOSUM62

Gapex 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents\_AA.\*

- 1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	264	4.2	545	US-08-936-107A-10	Sequence 10, Appl
2	171.5	2.7	10192	US-09-134-001C-3159	Sequence 3159, Ap
3	166.5	2.6	1015	US-08-537-210A-1	Sequence 1, Appl
4	166.5	2.6	1015	US-09-113-825-1	Sequence 1, Appl
5	166.5	2.6	2471	US-08-185-432-16	Sequence 16, Appl
6	166.5	2.6	2471	US-08-083-590A-19	Sequence 19, Appl
7	166.5	2.6	2471	US-08-532-384-19	Sequence 19, Appl
8	166.5	2.6	2471	US-08-899-232-1	Sequence 1, Appl
9	163.5	2.6	3135	US-08-323-170B-2	Sequence 2, Appl
10	163.5	2.6	3135	US-08-954-441-2	Sequence 2, Appl
11	153	2.4	1964	US-09-467-997-1	Sequence 1, Appl
12	145.5	2.3	1068	US-08-537-210A-2	Sequence 2, Appl
13	145.5	2.3	1068	US-09-113-825-2	Sequence 2, Appl
14	145.5	2.3	2556	US-08-185-432-17	Sequence 17, Appl
15	145.5	2.3	2556	US-08-083-590A-20	Sequence 20, Appl
16	145.5	2.3	2556	US-08-532-384-20	Sequence 20, Appl
17	145.5	2.3	2556	US-08-899-232-2	Sequence 2, Appl
18	143.5	2.3	2285	US-09-308-375-2	Sequence 2, Appl
19	141	2.2	1078	US-08-264-534-32	Sequence 32, Appl
20	141	2.2	1078	US-08-083-590A-11	Sequence 11, Appl
21	141	2.2	1078	US-08-465-500-32	Sequence 32, Appl
22	141	2.2	1078	US-08-346-128-32	Sequence 32, Appl
23	141	2.2	1078	US-08-532-384-11	Sequence 11, Appl
24	141	2.2	1078	US-08-893-828-32	Sequence 32, Appl
25	139	2.2	219	US-08-902-516-2	Sequence 2, Appl
26	138	2.2	1664	US-08-537-210A-3	Sequence 3, Appl
27	138	2.2	1064	US-09-113-825-3	Sequence 3, Appl

28	138	2.2	2523	1	US-08-185-432-18	Sequence 18, Appl
29	138	2.2	2523	4	US-08-899-232-3	Sequence 3, Appl
30	137	2.2	1388	4	US-09-572-191-2	Sequence 2, Appl
31	137	2.2	1388	4	US-09-723-262-2	Sequence 2, Appl
32	137	2.2	1388	4	US-09-723-219-2	Sequence 2, Appl
33	137	2.2	3696	4	US-09-134-001C-5080	Sequence 5080, Ap
34	136	2.1	878	4	US-09-134-001C-4578	Sequence 4378, Ap
35	134	2.1	1211	4	US-08-425-061-24	Sequence 4820, Ap
36	134	2.1	1852	1	US-08-825-886-24	Sequence 24, Appl
37	134	2.1	1852	1	US-08-425-061-16	Sequence 16, Appl
38	134	2.1	1863	1	US-08-480-784-2	Sequence 2, Appl
39	134	2.1	1863	1	US-08-483-553-2	Sequence 2, Appl
40	134	2.1	1863	1	US-08-487-002-2	Sequence 2, Appl
41	134	2.1	1863	1	US-08-483-554B-2	Sequence 2, Appl
42	134	2.1	1863	1	US-08-488-011B-2	Sequence 2, Appl
43	134	2.1	1863	2	US-08-825-886-16	Sequence 16, Appl
44	134	2.1	1863	4	US-08-850-727-2	Sequence 2, Appl
45	134	2.1	1863	4	US-08-850-727-2	Sequence 2, Appl

## ALIGNMENTS

RESULT 1  
US-08-936-107A-10  
; Sequence 10, Application US/08936107A  
; Patent No. 6403306  
; GENERAL INFORMATION:  
; APPLICANT: Stephens, David S.  
; APPLICANT: Swartley, John S.  
; TITLE OF INVENTION: Serogroup-Specific Nucleotide Sequences  
; TITLE OF INVENTION: in the Molecular Typing of Bacterial Isolates and the  
; TITLE OF INVENTION: Preparation of Vaccines Thereo  
; NUMBER OF SEQUENCES: 51  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: US  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE: 23-SEP-1997  
; APPLICATION NUMBER: US/08/936,107A  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/827,622  
; FILING DATE: 09-APR-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Caruthers, Jennie M.  
; REGISTRATION NUMBER: 34,464  
; REFERENCE/DOCKET NUMBER: 77-97  
; TELEPHONE: (303) 499-8080  
; TELEFAX: (303) 499-8089  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 545 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-936-107A-10

Query Match 4.2%; Score 264; DB 4; Length 545;  
Best Local Similarity 27.8%; Pred. No. 2.5e-13;  
Matches 90; Conservative 52; Mismatches 130; Indels 52; Gaps 13;

QY 135 LPANITLDSLPSYFHSASDIFNVAKPNPSNVVVFDSTKDVEDAHSLKLG-NS 193

Db 97 LPSNLTLK--PALCIILSHKEDFLN-----KFLLTISSENKLOQKFNQIKNPKS 145  
Qy 194 ROTWRGY-----LTTDKKEVPGVL-----LMQDLAFLSGFPPTFKETNQLKT 235  
Db 146 VNEIWTDLFSIAHVDKMLSTDRSLSSISQFWFRLECKEDKFI-----LFSTANRYSK 200  
Qy 236 KLPELSSKVKLLQLYSEASVALLKLNPKDFQELANKQTKKNTIDGKELTISPAYLLWD 295  
Db 201 KLWKH-----SIKNQLFKEG-----IRNYSEISLSPYEEDHNFIDIDLFTWVNSEDKNQ 251  
Qy 296 LSALSQSKQD---EDISARFEDNEELRYSLSRIERHAPWVRNFIWVTCQIPSWNLND 352  
Db 252 -ELYKTKPDNFNSDATSFSLRDELKFAIRSWEMSGSFIKIFIVSNCAPPAILDLNN 310  
Qy 353 PRVTIVHODVFNRLSHLTPSSPAIESHVHRIEGLSKQFIYLLNDVDFKQWMPDDFYS 412  
Db 311 PKIQWVHEEIMPO-SALPTFSSHAITSLSLHPIGISNFIYSNDDFLTKPLNKDNFFY 369  
Qy 413 HSKQKQVYL-TWVPVN-CAGGCP 433  
Db 370 SNGIAKRLLEAWGNVNGECTEGEP 393

RESULT 2

US-09-134-001C-3159  
; Sequence 3159, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134, 001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064, 964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055, 779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 3159  
; LENGTH: 10182  
; TYPE: PRN  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-3159

Query Match 2.78; Score 171.5; DB 4; Length 10182;  
Best Local Similarity 17.28; Pred. No. 0.0015;  
Matches 221; Conservative 212; Mismatches 461; Indels 389; Gaps 57;  
Qy 15 VPGSGDEQDPRIDGKLSRDQYH---VLFDSYRDNIAKSKSFQNRCL-L-PMPIDVYVT 70  
Db 7255 VPSSTLDKATIEDALLEQNARESLHGEQKLOEAKNAVA--EIDNLOALNPGQVLAET 7312  
Qy 71 WVN--GTDELLKELQVREMEBEOKAMREILGKNTTEPTTKSEKOLECLLTHCIVKM 128  
Db 7313 LVNQASTPEVQEAQAKA-LNEAMKALK-----TEINKBOIKADSRVNA----- 7359  
Qy 129 LVIDPALPANITKLDPLSLYSPHSASDIFNVAKPKNPSTNVVVFSTKQVEDAHSGL 188  
Db 7360 ---DSGLQANYN-----SALNYGSOIATTOP--PELNKD-VINRATQTIKTAENN- 7404  
Qy 189 LKGNRSQTWVRGYLTDTTKEVPLVMDLAFLSGFPPTFEK-----TNQKTK-----LP 238  
Db 7405 LNGQSK-----LAEAKSDG---NQSIEHLQGLTQSQKDKQHDNLNQACTKQOQVDDIV 7453  
Qy 239 ENL-----SSKVKLLQLYSEASVALLKLNPKDFQELANKQTKKNTIDGKELTISPAYLLW 294  
Db 7454 NNSKQLDNSMNOLOQIVVNDNT--VKQNSDFINEDSSQDAYNHAIOAAKDLITAHPTIM 7511  
Qy 295 DLSAISQSKQD-----EDISAR--FEDNEELRYSLSRIERHAPWVRNFIWVTCQIPSW 347  
Db 7512 DKNQIDQAIENITKQALNDLHGSNKLSEDKKASEOL-----QMLNSLTNGQ----- 7557

Qy 348 LNLNPRVTIYTHQDVFNRLSHLTPESSPA---IESHVHRIEGLSKQFIYLLNDV--- 399  
Db 7558 -----KDTIILNH-----IFSAPTSQYGEKIASAKQLNNTMKALRDSIADNNE 7600  
Qy 400 -----MEGKDVWPDFFYSH--SRGQKQVYLTWVPVPCAEGCPGSGWIKDGYCDKACNNSAC 451  
Db 7601 ILQSSKYFNESEQONAYNAQVNAKAKNIINDQPTPVMADEIQSVLNEVKTK----- 7653  
Qy 452 DWGDCGSGNSGGSRYIAGGGGTGSGVGPQWPGGGGINSYSCYNOGCANSWLAKDKCQD 511  
Db 7654 -----DNLHGQDKLANDKTDQ-----ATLNALNYLNQA-----Q 7693  
Qy 512 AGNVLSCFGDAGDCQDHFHELYKVIPLNQTHYIIPKGECLPYFSEFAEVAKRGVGAYS 571  
Db 7694 RGNL-----ETKQVNSNSRPEQKVQVQVLANQLNDAMK-----LDDALT 7722  
Qy 572 DNPIIRHASIAKWKTIHLIMHSGMNATTIHFNLTFQWNEDEEFKMQITVEVDTRGPKL 631  
Db 7723 GNDAIKQTS-----NYINEDTSQQVNFDEYDRG--- 7751  
Qy 632 NSTAQGYENLVSPITLLPEAEILFEDIPKEKRPKFRKRDVNSTRRRAQAEVKIPLVNIS 691  
Db 7752 -----KNIVA-----EOTNPMSPPTNINTIADKITAKNLDLHGQV 7796  
Qy 692 LLPKDAQLSLNTLDIQLHGDITLKGYNLSKLSALLRSFLMSQ-----HAKIKNQAIITD 746  
Db 7787 KLKQAOQOSINTIN-----QMTGLNQAOKEQLNQEIQTQTRSEVHVQVINKAQAALND 7838  
Qy 747 ETN---DSLVAPQEQKVHKSILPNSLG-----VSELRQLRTPPAVSKVNGHDOGNP 796  
Db 7839 SMNTLRQSTIDHEVQKTSNYINETVQNTAYNNAVDKVKOII-----NQTSP 7887  
Qy 797 ---PLDLE--TTAREFVE---THQKTIGNTVTKKPPSLIPLVESQMTKEKKITGKEKEN 849  
Db 7888 TWNPLEVERATSNVKISKDLHGERELNDN--KNSKTTAVNHLNDLNAQAKEALTHEIEQ 7945  
Qy 850 SRMEENAENHIGVTEVLIG--RKLOHYTDSYLGFLPWEKKYKFLDLDBEESLKTQLAYF 907  
Db 7946 ATIVSQVNNIYNKAKALNNDMKLK-----DIVAQDNVRSNNYI 7986  
Qy 908 TDSKNRARKRDTFFADSLRYVNVKILNSKEGFTSRKVPAPHMHIDRIYMQELODMFPEEF 967  
Db 7987 NEDSTPQNNYNDTINHAOSIIDQVANPT-----MSH--DEI----- 8020  
Qy 968 DKTSFKVHRHS-----EDMOFAFSFYFLMSAVQPLN-----ISOVFDEVDTDQSGV 1014  
Db 8021 -ENAINNIKHAINALDGEHKLQOAKENANLINSNLNDLNAPORDAINRLVNEAQTRKVA 8079  
Qy 1015 LSDREIRTLATRIHELPLSLQDLTGL--EHLMLINGSKMLPADITQNLNIPPTQESYSDPN 1072  
Db 8080 EQLOSQAALNDAMKHLRNSIQNSSVRQESKYINAS---DAKKEQYNHAREVENIINEQ 8136  
Qy 1073 LPVTVTSVNTCKPVTDKTHKAYKDKNKYFEIMG-----E 1108  
Db 8137 HPTLDKEII---KQLPDGVNQANNLDN--GVELLDADKQNAHQSIPTLHMLNQAOQNALN 8191  
Qy 1109 EETAFKMITNVSHVVGQ---LDDIRKNPRKEVCLNDNIDHNHDKAQTVKAVLRDPFYSM 1165  
Db 8192 EKINNATVTEVAALIGQAKLLDHAM-----ENLEESIKKEQVK-----QSS 8234  
Qy 1166 FPIPSQFELPREYRNFELMHHEL 1188  
Db 8235 NYINEDSDVOETYNDAVDHVTEI 8257

RESULT 3

US-08-537-210A-1  
; Sequence 1, Application US/08537210A  
; Patent No. 5780300  
; GENERAL INFORMATION:  
; APPLICANT: Artavanis-Tsakonas, Spyridon  
; APPLICANT: Fortini, Mark

; APPLICANT: Matsuno, Kenji  
 ; TITLE OF INVENTION: MANIPULATION OF NON-TERMINALLY  
 ; NUMBER OF SEQUENCES: 4  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Pennie & Edmonds  
 ; STREET: 1155 Avenue of the Americas  
 ; STATE: NY  
 ; COUNTRY: USA  
 ; ZIP: 10036/2711  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSeq Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/537,210A  
 ; FILING DATE: 29-SEP-1995  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Mistrock, S. Leslie  
 ; REGISTRATION NUMBER: 18,872  
 ; REFERENCE/DOCKET NUMBER: 7326-027  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 212-790-9090  
 ; TELEX: 66141 PENNIE  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1015 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS:  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: protein  
 ; FEATURE:  
 ; NAME/KEY: hum N (Human No. 5780300ch 2)  
 ; LOCATION: 1155...2169  
 ; OTHER INFORMATION: Highly conserved ankyrin repeat  
 ; OTHER INFORMATION: region of No. 5780300ch  
 ; US-08-537-210A-1

Query Match 2.6%; Score 166.5; DB 1; Length 1015;  
 Best Local Similarity 19.7%; Pred. No. 9.1e-05;  
 Matches 110; Conservative 64; Mismatches 176; Indels 207; Gaps 26;  
 QY 424 PVPNCAECPGSKWIKDGYCDKACNNSACDWDGDCSGNSGSRVIAGGGTGTGVCQPW 483  
 DB 267 PPATCLSOYCADKARDGVCDACNSHACQWDGDC-----SLTMENPW 309  
 QY 484 -QFGGINSVSYCQGC-----ANS-----WLDKPF-----CDQACN 514  
 DB 310 ANCSPLPCWDYINNQCDELNTVECLDNFECGNSKTKYDKYCADHFKDNHCNOCN 369  
 QY 515 VLSGFGDAGCGDGHFELHYK-----VILLPNQTHYIIPKGECLPYFSFAEV----- 561  
 DB 370 SEECGWGLDCAADQPENLAEGTLIVVLMPE-----PEQLQDARSFLRGLTLLHTN 422  
 QY 562 --AKRGVEGAYSDNPIIRHASIAKWKTIHLMHSGNATTIHNLFTFQNTDEEFK--- 616  
 DB 423 LRIKRDQSGELWYYPYGEKSAAMKKQ-----RWRSLPGEQGEVA 465  
 QY 617 -MQITVEVDYREGKLNSTAKGYENLVSPITLLPRAEILFDPKRPFKRHDVNS 675  
 DB 466 GSKVFLEIDNRQ---CVQDSDHCFKNTDAAALLAS-----HAIQG 503  
 QY 676 TRRAQEEVKIPLVNI---SLLPKDAQLSLNTLDLQ-----EUGDIT 714  
 DB 504 T-----LSPLVSVSVSESLTPERTQL-LYLLAVAVVILFIILLGVIMAKRRKHGSLW 556  
 QY 715 L-KGYNLKSAIILRSFLMNSQHAQ-----IKNOAITDITDNDLSLVAPOEKQVHKS 763  
 DB 557 LPEGFTLRDRA-----SNHKKRREPVGODAVGLANLSVQVSEAN--LI----- 596

QY 764 ILPNSLGVSRLORLTTPPAVSVKVNHDQG-----QNPPLDLETTARFRVETHQKTIGGN 819  
 DB 597 ----GTGTSEHWVDDEGPQPK-KVKAEDALLSEDDPIDR-----RPWTQQLHEAA 643  
 QY 820 VTREKPPSLVPLESOMTKB-----KKITGKEKENSMEENAEN---H 859  
 DB 644 DIRTPSLALTTPQAEQEVLDVNVNRPDGCPTPLMLASLRGGSSDLSDDEDAEDSSAN 703  
 QY 860 IGVTEVLLGRKLQHYTD 876  
 DB 704 IITDLVYQASLQAQTD 720

RESULT 4  
 US-09-113-825-1  
 ; Sequence 1, Application US/09113825  
 ; Patent No. 6149902  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Attavanis-Tsakonas, Spyridon  
 ; APPLICANT: Fortini, Mark  
 ; APPLICANT: Matsuno, Kenji  
 ; TITLE OF INVENTION: MANIPULATION OF NON-TERMINALLY  
 ; TITLE OF INVENTION: DIFFERENTIATED CELLS USING THE NOTCH PATHWAY  
 ; NUMBER OF SEQUENCES: 4  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Pennie & Edmonds  
 ; STREET: 1155 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: NY  
 ; COUNTRY: USA  
 ; ZIP: 10036/2711  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSeq Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/113,825  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/537,210  
 ; FILING DATE: 29-SEP-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Mistrock, S. Leslie  
 ; REGISTRATION NUMBER: 18,872  
 ; REFERENCE/DOCKET NUMBER: 7326-027  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 212-790-9090  
 ; TELEX: 66141 PENNIE  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1015 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS:  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: protein  
 ; FEATURE:  
 ; NAME/KEY: hum N (Human No. 6149902ch 2)  
 ; LOCATION: 1155...2169  
 ; OTHER INFORMATION: Highly conserved ankyrin repeat  
 ; OTHER INFORMATION: region of No. 6149902ch  
 ; US-09-113-825-1

Query Match 2.6%; Score 166.5; DB 4; Length 1015;  
 Best Local Similarity 19.7%; Pred. No. 9.1e-05;  
 Matches 110; Conservative 64; Mismatches 176; Indels 207; Gaps 26;  
 QY 424 PVPNCAECPGSKWIKDGYCDKACNNSACDWDGDCSGNSGSRVIAGGGTGTGVCQPW 483  
 DB 267 PPATCLSOYCADKARDGVCDACNSHACQWDGDC-----SLTMENPW 309

TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2471 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-185-432-16

Query Match 2.6%; Score 166.5; DB 1; Length 2471;  
Best Local Similarity 19.7%; Pred. No. 0.00039;  
Matches 110; Conservative 64; Mismatches 176; Indels 207; Gaps 26;

QY 424 PYPNCAEGCPGSGWIKGYCDKACNNSACDWDGDCSGNSGSGRYTAGGGTSGISGVGPW 483  
DB 1421 PPATCLSQYCADKARDGVCDACNSHACWDGDC-----SLTMENPW 1463  
QY 484 -QFGGINSVSYCNOGC-----ANS-----WLDKF-----CDQACN 514  
DB 1464 ANCSPLPCWDYINNQCDELNTVECLFDNFEQCGNSKTKYKCYADHFKDNHCNOGCN 1523  
QY 515 VLSGFDAGDCQDHFHLYK-----VILLPNQTHYIIPKGECLPYFSAEV-----561  
DB 1524 SEECGWDGLDCAADOPENLAEGTLVIVLMP-----PEQLQDARSFLRALGTLHTN 1576  
QY 562 --AKRGVEGAYSDNPIIRHASIANKWKTIHLIMHSGMNATTIHFNLTFQNTDEEFK---616  
DB 1577 LRIKRDQSGLWVYYPYGEKSAAMKQ-----RMTRSLPGEQEVEA 1619  
QY 617 -MQITVEVDTRGPKLNSTAQGYENLVSPITLLPEAEILFEDIPEKRFKFRHDVNS 675  
DB 1620 GSKVFEIDNRQ---CVQDSHCFKNTDAAALAS-----HAIOG 1657  
QY 676 TRAQEEVKIPLVNI---SLLPKDAQLSLNTLDLQ-----EHGDTT 714  
DB 1658 T-----LSYPLVSVVSESLTPERTQL-LYLLAVAVIILLGIVIMAKRKHGSLW 1710  
QY 715 L-KGYNLKSALLRSLFNMNSOHAKE-----TKNQAITDETNDLSIVAPQEKOVHKS 763  
DB 1711 LPEGFTLRDA-----SNHRRPVGQDAVGLKNLSVQVSEAN--LI-----1750  
QY 764 ILPNSLGVSERLQRLTFPAVSVKVNHGDOG---QNPPLDLETTARFRVETHTKTIGN 819  
DB 1751 ---GTGTSEHWVDDEGPQPK-KVKAEDAEALLSEEDPIDR-----RPWTQOHLAAA 1797  
QY 820 VTKEKPPSLIVPLESOMTKE-----KKITGKEKENSMEENAEEN--H 859  
DB 1798 DIRTPSLALTPQAEQEVLDVNVVRGPDGCTPLMLASLRGSSDLSDEDEDAEDSSAN 1857  
QY 860 IGVTEVLLGRKLQHYTD 876  
DB 1858 IITDLVYQASLQAQTD 1874

RESULT 6  
US-08-083-590A-19  
Sequence 19, Application US/08083590A  
Patent No. 5786158  
GENERAL INFORMATION:  
APPLICANT: Artavanis-Tsakonas, S. et al.  
TITLE OF INVENTION: Therapeutic And Diagnostic Methods  
TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And  
TITLE OF INVENTION: Nucleic Acids  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:

QY 484 -QFGGINSVSYCNOGC-----ANS-----WLDKF-----CDQACN 514  
DB 310 ANCSPLPCWDYINNQCDELNTVECLFDNFEQCGNSKTKYKCYADHFKDNHCNOGCN 369  
QY 515 VLSGFDAGDCQDHFHLYK-----VILLPNQTHYIIPKGECLPYFSAEV-----561  
DB 370 SEECGWDGLDCAADOPENLAEGTLVIVLMP-----PEQLQDARSFLRALGTLHTN 422  
QY 562 --AKRGVEGAYSDNPIIRHASIANKWKTIHLIMHSGMNATTIHFNLTFQNTDEEFK---616  
DB 423 LRIKRDQSGLWVYYPYGEKSAAMKQ-----RMTRSLPGEQEVEA 465  
QY 617 -MQITVEVDTRGPKLNSTAQGYENLVSPITLLPEAEILFEDIPEKRFKFRHDVNS 675  
DB 466 GSKVFEIDNRQ---CVQDSHCFKNTDAAALAS-----HAIOG 503  
QY 676 TRAQEEVKIPLVNI---SLLPKDAQLSLNTLDLQ-----EHGDTT 714  
DB 504 T-----LSYPLVSVVSESLTPERTQL-LYLLAVAVIILLGIVIMAKRKHGSLW 556  
QY 715 L-KGYNLKSALLRSLFNMNSOHAKE-----TKNQAITDETNDLSIVAPQEKOVHKS 763  
DB 557 LPEGFTLRDA-----SNHRRPVGQDAVGLKNLSVQVSEAN--LI-----596  
QY 764 ILPNSLGVSERLQRLTFPAVSVKVNHGDOG---QNPPLDLETTARFRVETHTKTIGN 819  
DB 597 ---GTGTSEHWVDDEGPQPK-KVKAEDAEALLSEEDPIDR-----RPWTQOHLAAA 643  
QY 820 VTKEKPPSLIVPLESOMTKE-----KKITGKEKENSMEENAEEN--H 859  
DB 644 DIRTPSLALTPQAEQEVLDVNVVRGPDGCTPLMLASLRGSSDLSDEDEDAEDSSAN 703  
QY 860 IGVTEVLLGRKLQHYTD 876  
DB 704 IITDLVYQASLQAQTD 720



```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/083,590A
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 8698864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2471 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-083-590A-19

Query Match          2.6%; Score 166.5; DB 1; Length 2471;
Best Local Similarity 19.7%; Pred. No. 0.00039;
Matches 110; Conservative 64; Mismatches 176; Indels 207; Gaps 26;

QY 424 PVPNCABGCPGSKIWDYCDKACNNSACDWDGDCSGNSGSGRYIAGGGTSGIGVGPW 483
Db 1421 PPATCLSQYCADKARDGVCDACNSHACQWDGDC-----SLTMENPW 1463
QY 484 -QFGGINSVSYCNQGC-----ANS-----WLADKF-----CDQACN 514
Db 1464 ANCSPLPCWDYINNOCDCLNTVECLFDFNFCQGNSTCKYDKYCADHFKDNHCNQC 1523
QY 515 VLSCGFDAGDCQDHFHELYK-----VILLPNQTHYIIPKGECLPYFSFAEV-----561
Db 1524 SEECGWGLDCAADQENLAEGTLVIVLMP-----PEQLQDARSFLRALGTLHTN 1576
QY 562 --AKRGVEGAYSDNPIIRHASIANKWKTIHLIMHSGMNTTIHFNLTFQNTNDEEFK---616
Db 1577 LRIKRSOGELMVYPYGEKSAAMKQ-----RMTRRSLPGEQOEVA 1619
QY 617 -MQITVEVDTRGPKLNSTAQGYENLYSPITLLPEAEILFEDIPKEKRPFKRHDVNS 675
Db 1620 GSKVFEIDNRQ---CVQDSHCHFKNTDAAALLAS-----HAIQG 1657
QY 676 TRAQEEVKIPLVNI---SLLPKDAQLSLNTLDLQL-----EHGDT 714
Db 1658 T-----LSYPLSVSVSESLTPERTQL-LYLLAVAVVILFILLGVIMAKRRKHGSLW 1710
QY 715 L-KGYNLKSALLRSLFMSQAHK-----IKNOAITDETNDLSVAPQEKQVHKS 763
Db 1711 LPEGETLRDA-----SNHKREPVGDVAGLKNLSVQVSEAN--LI-----1750
QY 764 ILPNSLGSYERLQRLTFRAVSVKVNHDQ-----QNPLDLLETTARFVTHTKTCGN 819
Db 1751 ----GTGTSSEHWVDEGQPK-KYKAEDALLSEDDPIDR-----RPTQOHLAA 1797
QY 820 VTKEPPSLVPLESOMTKE-----KKITGKEKENSMEENAEN---H 859
Db 1798 DIRRTPSLALTPQAEQEVLDVNVNRPDGTPLMLASLRGGSDLSDEDEDAEDSSAN 1857
QY 860 IGVTEVLLGRKLOHYTD 876
Db 1858 IITDLVYOGASLQAQTD 1874

RESULT 7
US-08-532-384-19
; Sequence 19, Application US/08532384
```

```
; Patent No. 6083904
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, S. et al.
; TITLE OF INVENTION: Therapeutic And Diagnostic Methods
; TITLE OF INVENTION: And Compositions Based On No. 6083904ch Proteins And
; TITLE OF INVENTION: Nucleic Acids
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/532,384
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/083,590
; FILING DATE: 25-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 8698864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2471 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-532-384-19
```

```
Query Match          2.6%; Score 166.5; DB 3; Length 2471;
Best Local Similarity 19.7%; Pred. No. 0.00039;
Matches 110; Conservative 64; Mismatches 176; Indels 207; Gaps 26;

QY 424 PVPNCABGCPGSKIWDYCDKACNNSACDWDGDCSGNSGSGRYIAGGGTSGIGVGPW 483
Db 1421 PPATCLSQYCADKARDGVCDACNSHACQWDGDC-----SLTMENPW 1463
QY 484 -QFGGINSVSYCNQGC-----ANS-----WLADKF-----CDQACN 514
Db 1464 ANCSPLPCWDYINNOCDCLNTVECLFDFNFCQGNSTCKYDKYCADHFKDNHCNQC 1523
QY 515 VLSCGFDAGDCQDHFHELYK-----VILLPNQTHYIIPKGECLPYFSFAEV-----561
Db 1524 SEECGWGLDCAADQENLAEGTLVIVLMP-----PEQLQDARSFLRALGTLHTN 1576
QY 562 --AKRGVEGAYSDNPIIRHASIANKWKTIHLIMHSGMNTTIHFNLTFQNTNDEEFK---616
Db 1577 LRIKRSOGELMVYPYGEKSAAMKQ-----RMTRRSLPGEQOEVA 1619
QY 617 -MQITVEVDTRGPKLNSTAQGYENLYSPITLLPEAEILFEDIPKEKRPFKRHDVNS 675
Db 1620 GSKVFEIDNRQ---CVQDSHCHFKNTDAAALLAS-----HAIQG 1657
QY 676 TRAQEEVKIPLVNI---SLLPKDAQLSLNTLDLQL-----EHGDT 714
Db 1658 T-----LSYPLSVSVSESLTPERTQL-LYLLAVAVVILFILLGVIMAKRRKHGSLW 1710
QY 715 L-KGYNLKSALLRSLFMSQAHK-----IKNOAITDETNDLSVAPQEKQVHKS 763
```

Db 1711 LPEGFTLRDA-----SNHKREPVGQDAVGLKNSLVQVSEAN--LI-----1750  
QY 764 ILPNSLGVSERLQRLTFPAVSKVNGHDG---QNPPLDLETTARFRVETHQKTIGGN 819  
Db 1751 ----GTGISEHWVDEGQPK-KVKAEDALLSEDDPIDR-----RPWTQOQHLEAA 1797  
QY 820 VTKEPPSLIVPLESOMTKE-----KKITGKEKENSMEENAEAN--H 859  
Db 1798 DIRRTPSLALTPPAEQEVVDVNVVRPGDCTPLMLASLRGGSSDLSDEDAEDSSAN 1857  
QY 860 IGVTEVLLGRKLQHYTD 876  
Db 1858 IITDLVYQASLQAQTD 1874  
RESULT 8  
US-08-899-232-1  
; Sequence 1, Application US/08899232  
; Patent No. 6436650  
; GENERAL INFORMATION:  
; APPLICANT: Artavanis-Tsakonas, Spyridon  
; APPLICANT: Qi, Hulin  
; TITLE OF INVENTION: ACTIVATED FORMS OF NOTCH AND METHODS BASED THEREON  
; FILE REFERENCE: 7326-046  
; CURRENT APPLICATION NUMBER: US/08/899,232  
; CURRENT FILING DATE: 1997-07-23  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 2471  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-08-899-232-1

Query Match 2.6%; Score 166.5; DB 4; Length 2471;  
Best Local Similarity 19.7%; Pred. No. 0.00039;  
Matches 110; Conservative 64; Mismatches 176; Indels 207; Gaps 26;  
QY 424 PVPNCAECPGCSWIKDGYCDKACNNSACDWDGDCSGNSGSRVYAGGGTSGVGPW 483  
Db 1421 PPATCLSOYCADKARDGVCDEACNSHACQWDGDC-----SLTMENPW 1463  
QY 484 -QFGGINSVSYCNQC-----ANS-----WLADKF-----CDQACN 514  
Db 1464 ANCSPLPCWDYINNQCDELNCTVECLDFNFCQGNSTCKYDKYCADHFKNHCNCGN 1523  
QY 515 VLSGCFDAGDCGQHFHELYK-----VILLPNQTHYIIPKGECLPYFSFAV-----561  
Db 1524 SBECDGDLCDAAQDQENLAEGTLIVVILMP-----PEQLQDARSFLRALGTLHTN 1576  
QY 562 --AKRGVEGAYSDNPIIRHASTANKWKTIIHMHSGMNATTIHFNLTFONTNDEEFK---616  
Db 1577 LRIKDSGELMVYPIYCEKAAMKQ-----RMTFRSLPGQEQEVA 1619  
QY 617 -MQITVEDTREGPKLNSTAGQENLVSPITLLPEAEILFDEIPKEKRPFKFRHDVNS 675  
Db 1620 GSKVFEIDNRQ---CVQDSBHCFKNTDAAAALLAS-----HAIOG 1657  
QY 676 TRAAEEVKIPLVNI---SLLPKDAQLSLNTLDLQL-----EHGDIT 714  
Db 1658 T-----LSYPLVSVVSESLPTEPTQL-LYLLAVAVIILFIILGVIMAKRKRKHSGLW 1710  
QY 715 L-KGYNLKSALLRSFLMNSQAK-----IKNQAITDETNTSLVAPQEQVHKS 763  
Db 1711 LPEGFTLRDA-----SNHKREPVGQDAVGLKNSLVQVSEAN--LI-----1750  
QY 764 ILPNSLGVSERLQRLTFPAVSKVNGHDG---QNPPLDLETTARFRVETHQKTIGGN 819  
Db 1751 ----GTGISEHWVDEGQPK-KVKAEDALLSEDDPIDR-----RPWTQOQHLEAA 1797  
QY 820 VTKEPPSLIVPLESOMTKE-----KKITGKEKENSMEENAEAN--H 859  
Db 1798 DIRRTPSLALTPPAEQEVVDVNVVRPGDCTPLMLASLRGGSSDLSDEDAEDSSAN 1857

QY 860 IGVTEVLLGRKLQHYTD 876  
Db 1858 IITDLVYQASLQAQTD 1874

RESULT 9  
US-08-323-170B-2  
; Sequence 2, Application US/08323170B  
; Patent No. 5733772  
; GENERAL INFORMATION:  
; APPLICANT: Williamson, Kim C.  
; APPLICANT: Kaslow, David C.  
; TITLE OF INVENTION: Cloning and Expression of Plasmodium  
; TITLE OF INVENTION: falciiparum Transmission-Blocking Target Antigen, Pfs230  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, 8th Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/323,170B  
; FILING DATE: 13-OCT-1994  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/010,409  
; FILING DATE: 29-JAN-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Quine, Jonathan A.  
; REGISTRATION NUMBER: P-41,261  
; REFERENCE/DOCKET NUMBER: 015280-113100US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3135 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-323-170B-2

Query Match 2.6%; Score 163.5; DB 1; Length 3135;  
Best Local Similarity 19.7%; Pred. No. 0.001;  
Matches 206; Conservative 124; Mismatches 378; Indels 337; Gaps 52;

QY 236 KLPENISSKVKLLQLYS-EASVALLKLNPK-----DFQLN-----KQPKK 276  
Db 1654 QLEEVYNDIESLELDIEQYVLQVNLKAPKMLMSAQIHNNRHVCDFSKNLIVPELKK 1713  
QY 277 NMTIDCKELTISPAYLLWLSAI-----SOSKODEDISASREFDNE---ELR-----320  
Db 1714 KEELGGNPVNIHCYALLKPLDLYVKCPTSKDYEAAKNISENDNEYELQVLSIEKRF 1773  
QY 321 YLSRISERHAPVVRNIFVITNGQIPSWLNLNDPRVTIVTHQDVRNLSHLPTFSSPAIES 380  
Db 1774 HNFETLESKKPGNGDV-VVHNGVVDTPGLDNS-----TFEYFKNIKIKP-----1818  
QY 381 HVIRIEGLSKFIYLLNDVWFGKD---VWPDDEYFHSKQK-----VLTWPVPCARG 431  
Db 1819 -----DKFFEKVINEYDDTEEEKDLESILPGAIVSPMKVILKKKDPPTSTAAAFVPPVIVP- 1872  
QY 432 CPGSWTIKDGYCDKACNNSACDWDGDCSGNSGSRVYAGGGTSGVGPQ-----WQF 485  
Db 1873 -----KDLHFKECNTE-----YKDNQYISYNGIITHIDISNKRKINGCDF 1916

QY 486 GGINSV-----SYNQCAGNSWLADKFCDOACNVLSGCGFDAGCGQDHFL 533  
Db 1917 STNNSILTSVVKLVNGETKNCENINNNNEVFGIICDNETNL-----DPEKC---FHEI 1967  
QY 534 YKVIILPNOT---HYIIPKGCPLPYFSAEVAKRGVEGAYSDNPI--IRHASIANKWT 587  
Db 1968 YS---KDNKTVKKFEVIPN---IDIFSLHNSNKKV--AYAKVPLDYINKLLFSCSCKT 2019  
QY 588 IHLIMHSGMNATTIHFNLTFQNTNDEEFKMQITVEVDREGPKLNSTAQKGYENLVSPIT 647  
Db 2020 SHNTIGTKMVT-----LNKDEKEEDFK-----TAQIKHNNVHLCN 2057  
QY 648 LLPEAEILFEDIPKEKRPKFRKRDVNSTRAQAEVKIPLVNIISLLPKDAQISLNTLDLQ 707  
Db 2058 FFDNPELTFDN-----NKIVLCKI-----DAEL-FSEVIQ 2087  
QY 708 L-----EHGDTILKGYNLSKSALLRFLMNSQAK--IKNOAIITDETNDLSVAPOE 757  
Db 2088 LPFQTKNVEEG-----VONEEYKFKSLKPSLVFDDNNDIKVIGKE 2129  
QY 758 KQVHKSILPNSIGVSRERLQRTFFPAVSVKVNCHDQGN-----PPLDLETTARFRV-ET- 810  
Db 2130 K-----NEVSIISLALGVYGNRIFTDKNKGKGGISFPFIPKODTDLKFIINETI 2181  
QY 811 -----HTQRTIGGNVTK-----EKPPSLIVPLESQMTKEKITGKEKENSME 853  
Db 2182 DNSNIQKRGIIYFVRKNVSENFKLCDFTTGSTSMLNLSQV-KEKCTVKKIKKGDIFG 2240  
QY 854 ENAEHIGV-----TEVLLGRKLOHYTDS-YLGLFLPWEKKY-----FLDLDEESL 900  
Db 2241 LKCPKGAIFPQACFSNVLYEYSDYEDSEHINYIHKDKKYNLKPDKVIELMDEN--- 2297  
QY 901 KTQLAYFDSKNRARKYR-DFFADSLRYVN-KILNSREGFTSRKVPAR--WPHMIDRIVM 956  
Db 2298 -----FRELNIQOYTGISNITDVLHFKNENLGNLPLNFKNHSTAYAKVPDFTFSIIN 2351  
QY 957 QELQDMFPE-----EFDKTSFHKVRHSEDMQAF-----SYFYVILMSAVQPLNIS 1001  
Db 2352 FSCCYNPEKHVGTQMVESNRNFNKNENYVKNFLNPKENIYKALLLDDERQKKIK 2411  
QY 1002 QVDFEVDVDSGLSDREIRLATRIHELPLSLQDLTGLEHMLI----- 1045  
Db 2412 QOQEEQOQEQ--ILKQDDR-----LSRHDDYKNKHTYILYDSNEHICDYKNES 2459  
QY 1046 -----NGSKMLPADITQLN-----NIPPTQESYDNPPLPVTKSLVYTC 1084  
Db 2460 LISTLPNDTKKIQKSICKINAKALDVVTIKCPHTKNTPKD---YFFN-----SSLTND 2511  
QY 1085 KPVTDKIHK-----AYKDNKRYRFEI 1105  
Db 2512 KKVITFDKKNFVYIDPTKKTFSL 2536

RESULT 10  
US-08-954-441-2  
; Sequence 2, Application US/08954441  
; Patent No. 6316000  
; GENERAL INFORMATION:  
; APPLICANT: Williamson, Kim C.  
; APPLICANT: Kaslow, David C.  
; TITLE OF INVENTION: Cloning and Expression of Plasmodium  
; TITLE OF INVENTION: falciparum Transmission-Blocking Target Antigen, Pfs230  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/954,441  
FILING DATE: 20-OCT-1997  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/323,170  
FILING DATE: 13-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/010,409  
FILING DATE: 23-JAN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Einhorn, Gregory P.  
REGISTRATION NUMBER: 38,440  
REFERENCE/DOCKET NUMBER: 015280-113110US  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3135 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-954-441-2

Query Match 2.6%; Score 163.5; DB 4; Length 3135;  
Best Local Similarity 19.7%; Pred. No. 0.001;  
Matches 206; Conservative 124; Mismatches 378; Indels 337; Gaps 52;  
QY 236 KLPENSSKVKLLQLYS-EASVALLKLNPK-----DFOELN-----KQTKK 276  
Db 1654 OLEEVYNDIESLELDKIEQYVLQVNLKAPKLMASQIHNRRVCDFSKNLLIVPESLKK 1713  
QY 277 NMTIDCKELTISPAYLLWDLISAI-----SOSKODEDISASRFEDNE---ELR----- 320  
Db 1714 KEELGSPVNIHCYALLKPLDLYVKCPTSKONYEAKVNISENDNEYELVISLEKRF 1773  
QY 321 YSLRSTERNHAPWVRNFIYVINGQIPSWLNLDNPRVITVHQDVFRNLHLPTFFSSPAIES 380  
Db 1774 HNPETLESKKPGNGDV-VVHNGVVDTPGVLDS-----TFEYFKNIKIKP----- 1818  
QY 381 HVHRIEGLSQKFIYLVNDVYMFQKD---VWPDDEFYSKSKQK-----VLTWPVPCABG 431  
Db 1819 -----DKFFEKVINEYDDTTEEKDLESILFGAIVSPMKVLKKKDPFTSYAAEFVPIVP- 1872  
QY 432 CPGSWIKDGYCDKACNNSACDWDGDCSGSGSRVYAGGCTGSGVGP-----WQF 485  
Db 1873 -----KDLHFKEVCNTE-----YKDENVYISYNGLIHIDISNKRKINGCDF 1916  
QY 486 GGINSV-----SYNQCAGNSWLADKFCDOACNVLSGCGFDAGCGQDHFL 533  
Db 1917 STNNSILTSVVKLVNGETKNCENINNNNEVFGIICDNETNL-----DPEKC---FHEI 1967  
QY 534 YKVIILPNOT---HYIIPKGCPLPYFSAEVAKRGVEGAYSDNPI--IRHASIANKWT 587  
Db 1968 YS---KDNKTVKKFEVIPN---IDIFSLHNSNKKV--AYAKVPLDYINKLLFSCSCKT 2019  
QY 588 IHLIMHSGMNATTIHFNLTFQNTNDEEFKMQITVEVDREGPKLNSTAQKGYENLVSPIT 647  
Db 2020 SHNTIGTKMVT-----LNKDEKEEDFK-----TAQIKHNNVHLCN 2057  
QY 648 LLPEAEILFEDIPKEKRPKFRKRDVNSTRAQAEVKIPLVNIISLLPKDAQISLNTLDLQ 707  
Db 2058 FFDNPELTFDN-----NKIVLCKI-----DAEL-FSEVIQ 2087  
QY 708 L-----EHGDTILKGYNLSKSALLRFLMNSQAK--IKNOAIITDETNDLSVAPOE 757  
Db 2088 LPFQTKNVEEG-----VONEEYKFKSLKPSLVFDDNNDIKVIGKE 2129  
QY 758 KQVHKSILPNSIGVSRERLQRTFFPAVSVKVNCHDQGN-----PPLDLETTARFRV-ET- 810

Db 2130 K-----NEVSISLAKGVYGNRIFFDKNGKKGEGISFFPIPKQDTDLKFIINETI 2181  
Qy 811 -----HTOKTIGNVTK-----BKPPSLVPLESOMTKKKITGKEKNSRME 853  
Db 2182 DNSNIQORGLIYIFVRKNVSENSEFKLDDFTTGSTSLMELNSQV-KEBKCTVKIKGDIIF 2240  
Qy 854 ENAENHIGV-----TSLVLRKLOHVTDS-YLGFPLPWEKKY-----FLDLLDEESL 900  
Db 2241 LKCPKFAIFPQAFCSNVLLYEYKSDYSEHINYYIHKDKYNLKPDKVIELMDEN--- 2297  
Qy 901 KTOLAYFTDSKNRARKR-DTFADSLRYVN-KILNSKGFSTSRKVPAAH-MPHMIDRIYV 956  
Db 2298 -----FRELQNIQYTGISNITDVLHFKFNGLNPLNFKNHYSTAYAKVPDTFNSIIN 2351  
Qy 957 QELQDMEPE-----EPDKTSFKVHRSEDMQFAP-----SYFYILMSAVQPLNIS 1001  
Db 2352 FSCNCPNPKHVYTMQVESDNRFNDIKNKNVKNFLNPNIEKYALLDDEERQKKIK 2411  
Qy 1002 QVFEVDTDSGVLSRDREIRTLATRIHELPLSLQDLTGLEHMLI----- 1045  
Db 2412 QOEEEOQEQ-----ILKQDDR-----LSRHDDYNKNHTYILYDSNEHICDYEKNES 2459  
Qy 1046 -----NCSKMLPADITQLN-----NIPPTQESYDYPNLPVTKSLVTNC 1084  
Db 2460 LISTLPNDTKKIQKSICKINAKALDVTIKCPHTKNFTPKD---YFPN-----SSLITND 2511  
Qy 1085 KPVTDKTHK-----AYKOKNKYRFEI 1105  
Db 2512 KKIIVTFDKKNFVYIIDPTKKTFSL 2536

## RESULT 11

US-09-467-997-1  
; Sequence 1, Application US/09467997  
; Patent No. 6379925

; GENERAL INFORMATION:  
; APPLICANT: Kitajewski, Jan

; APPLICANT: Uytendaele, Hendrik  
; TITLE OF INVENTION: ANGIOGENIC MODULATION BY NOTCH SIGNAL TRANSDUCTION

; FILE REFERENCE: 53863-A-PCT-US

; CURRENT APPLICATION NUMBER: US/09/467,997

; CURRENT FILING DATE: 1999-12-20

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1

; LENGTH: 1964

; TYPE: PRT

; ORGANISM: mouse

US-09-467-997-1

Query Match 2.4%; Score 153; DB 4; Length 1964;  
Best Local Similarity 28.1%; Pred. No. 0.0036;  
Matches 41; Conservative 11; Mismatches 32; Indels 62; Gaps 8;

Qy 424 PVPNC-----AEGCPESWTKDGYCDKACNNSACDWDGDCSGNSGGSYIAGGGTGSIGV 479  
Db 1162 PPRCORPGASGCEBGRG-GDGTCDAGCGSGPGGDWDGDC-----SLGV 1203

Qy 480 GPPWQFGGINSVVC-----NQCA-----NSWLADKF- 508

Db 1204 PDPWK---GCPPHSOWLLFRDGRCHQCDSEECFLDGYDCEIPTPTCIPAYDQICRDFH 1260

Qy 509 ---CDQACNVLSGCFDAGDC---GQD 528

Db 1261 NGHCEKGCNNAECGWDGDCRPEGED 1286

## RESULT 12

US-08-537-210A-2

; Sequence 2, Application US/08537210A

; Patent No. 5780300

; GENERAL INFORMATION:

; APPLICANT: Artavanis-Tsakonas, Spyridon  
; APPLICANT: Fortini, Mark  
; APPLICANT: Matsuno, Kenji  
; TITLE OF INVENTION: MANIPULATION OF NON-TERMINALLY  
; TITLE OF INVENTION: DIFFERENTIATED CELLS USING THE NOTCH PATHWAY  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: NY

; COUNTRY: USA  
; ZIP: 10036/2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/537,210A  
; FILING DATE: 29-SEP-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mistock, S. Leslie  
; REGISTRATION NUMBER: 18,872  
; REFERENCE/DOCKET NUMBER: 7326-027  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-790-9090  
; TELEFAX: 212-869-8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1068 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; FEATURE:  
; NAME/KEY: Human N1 (TAN-1)  
; LOCATION: 1152...2219  
; OTHER INFORMATION: Highly conserved ankyrin repeat  
; OTHER INFORMATION: region of No. 5780300ch

US-08-537-210A-2

Query Match 2.3%; Score 145.5; DB 1; Length 1068;  
Best Local Similarity 23.7%; Pred. No. 0.0055;  
Matches 50; Conservative 13; Mismatches 57; Indels 91; Gaps 11;

Qy 425 VPCNACGCPGSIKDGCDKACNNSACDWDGDCSGNSGGSYIAGGGTGSIGVGPWQ 484  
Db 301 LPQCQEDAGNK-----VCSLOCNNHACGWDGDC-----SLNENDPWK 338

Qy 485 -----FGGINSVSYCNO-GC-----ANSLADKFCDO 511  
Db 339 NCTQSLQCNKYFSDG-HCDSQCSAGCLDFDGCQRAEQCNPLYDQYCKDHFSDGHCDO 397

Qy 512 ACNVLSGCFDAGCGODHFE-----LYKVILLP-----NOTHYIIPKGECL----- 553  
Db 398 GCNSACEWDGLDCA-EHVPERLAAGTLVVVYVIMPEQLRNSFSFELMELSRVLHTNVVF 456

Qy 554 -----PYESFAE-----VAKRGVEG 568

Db 457 KRDAHGQQMIFPYIGREELRKHPIKRAAEG 487

## RESULT 13

US-09-113-825-2

; Sequence 2, Application US/09113825

; Patent No. 6149902

; GENERAL INFORMATION:

; APPLICANT: Artavanis-Tsakonas, Spyridon

; APPLICANT: Fortini, Mark

; APPLICANT: Matsuno, Kenji

;; TITLE OF INVENTION: MANIPULATION OF NON-TERMINALLY  
;; TITLE OF INVENTION: DIFFERENTIATED CELLS USING THE NOTCH PATHWAY  
;; NUMBER OF SEQUENCES: 4  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Pennie & Edmonds  
;; STREET: 1155 Avenue of the Americas  
;; CITY: New York  
;; STATE: NY  
;; COUNTRY: USA  
;; ZIP: 10036/2711  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette  
;; COMPUTER: IBM Compatible  
;; OPERATING SYSTEM: DOS  
;; SOFTWARE: FastSeq Version 2.0  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/113,825  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/537,210  
;; FILING DATE: 29-SEP-1995  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Mistrock, S. Leslie  
;; REGISTRATION NUMBER: 18,872  
;; REFERENCE/DOCKET NUMBER: 7326-027  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 212-790-9090  
;; TELEFAX: 212-869-8864  
;; TELEX: 66141 PENNIE  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1068 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS:  
;; TOPOLOGY: unknown  
;; MOLECULE TYPE: protein  
;; FEATURE:  
;; NAME/KEY: Human N1 (TAN-1)  
;; LOCATION: 1152...2219  
;; OTHER INFORMATION: Highly conserved ankyrin repeat  
;; OTHER INFORMATION: region of No. 6149902ch  
;; US-09-113-825-2

Query Match 2.3%; Score 145.5; DB 4; Length 1068;  
Best Local Similarity 23.7%; Pred. No. 0.0055;  
Matches 50; Conservative 13; Mismatches 57; Indels 91; Gaps 11;  
QY 425 VPNAEGCPGSKWDKCYDKACNNSACDWDGDCSGNSGGSRYIAGGGTSGIGVGPWQ 484  
Db 301 LPEQEDAGNK-----VCSLQCNHACGWDGDC-----SLNFPDPMK 338  
QY 485 -----FGGINSVSYCNO-GC-----ANSWLADKFCDDQ 511  
Db 339 NCTQSLQCNKYFSDG-HCDSQCNACGLFDGDCQRAEGCNPPLYDQYCKDHFSDGHCDDQ 397  
QY 512 ACNVLSCGFDAGCGQDHFHE-----LYKVILLP-----NQTHYIIPKGECL----- 553  
Db 398 GCNSAECEDWGLDCA-EHVPERLAAGTLVVVLMPPPEQLRNSSFHFLWELSRVLHTNVVF 456  
QY 554 -----PYFSFAE-----VAKRGVEG 568  
Db 457 KRDAHQOQMIFPYGEEELRKHPIKRAAEG 487

RESULT 14  
US-08-185-432-17  
; Sequence 17, Application US/08185432  
; Patent No. 5750652  
; GENERAL INFORMATION:  
; APPLICANT: Artavanis-Tsakonas, Spyridon  
; APPLICANT: Busseau, Isabelle  
; APPLICANT: Diederich, Robert J.

;; APPLICANT: Xu, Tian  
;; APPLICANT: Matsuno, Kenji  
;; TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND  
;; TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS  
;; NUMBER OF SEQUENCES: 23  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: PENNIE & EDMONDS  
;; STREET: 1155 Avenue of the Americas  
;; CITY: New York  
;; STATE: New York  
;; COUNTRY: U.S.A.  
;; ZIP: 10036-2711  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/185,432  
;; FILING DATE: 21-JAN-1994  
;; CLASSIFICATION: 530  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Mistrock, S. Leslie  
;; REGISTRATION NUMBER: 18,872  
;; REFERENCE/DOCKET NUMBER: 7326-006  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (212) 790-9090  
;; TELEFAX: (212) 869-8864/9741  
;; TELEX: 66141 PENNIE  
;; INFORMATION FOR SEQ ID NO: 17:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 2556 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: unknown  
;; MOLECULE TYPE: protein  
;; US-08-185-432-17

Query Match 2.3%; Score 145.5; DB 1; Length 2556;  
Best Local Similarity 23.7%; Pred. No. 0.023;  
Matches 50; Conservative 13; Mismatches 57; Indels 91; Gaps 11;  
QY 425 VPNAEGCPGSKWDKCYDKACNNSACDWDGDCSGNSGGSRYIAGGGTSGIGVGPWQ 484  
Db 1452 LPEQEDAGNK-----VCSLQCNHACGWDGDC-----SLNFPDPMK 1489  
QY 485 -----FGGINSVSYCNO-GC-----ANSWLADKFCDDQ 511  
Db 1490 NCTQSLQCNKYFSDG-HCDSQCNACGLFDGDCQRAEGCNPPLYDQYCKDHFSDGHCDDQ 1548  
QY 512 ACNVLSCGFDAGCGQDHFHE-----LYKVILLP-----NQTHYIIPKGECL----- 553  
Db 1549 GCNSAECEDWGLDCA-EHVPERLAAGTLVVVLMPPPEQLRNSSFHFLWELSRVLHTNVVF 1607  
QY 554 -----PYFSFAE-----VAKRGVEG 568  
Db 1608 KRDAHQOQMIFPYGEEELRKHPIKRAAEG 1638

RESULT 15  
US-08-185-590A-20  
; Sequence 20, Application US/08083590A  
; Patent No. 5786158  
; GENERAL INFORMATION:  
; APPLICANT: Artavanis-Tsakonas, S. et al.  
; TITLE OF INVENTION: Therapeutic And Diagnostic Methods  
; TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York

Search completed: April 13, 2003, 03:29:21  
Job time : 63.9002 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2003, 01:34:46 ; Search time 66.7772 Seconds  
(without alignments)  
2392.545 Million cell updates/sec

Title: US-10-023-888-2

Perfect score: 6340

Sequence: 1 METDTLLVLLWLLVPGSTG.....NRFLLHMHLEQENRAYDKLK 1199

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_101002.\*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1893	29.9	367	22	Peptide #2930 enco
2	1893	29.9	367	22	Human brain expres
3	1893	29.9	367	22	Peptide #2891 enco
4	1893	29.9	367	22	Peptide #2989 enco
5	1893	29.9	367	22	Peptide #2868 enco
6	1893	29.9	367	23	Human peptide enco
7	854.5	13.5	652	22	Drosophila melanog
8	460	7.3	132	22	Human protein sequ
9	352	5.6	68	22	Peptide #5693 enco
10	352	5.6	68	22	Human brain expres

11	352	5.6	68	22	AAW71337	Human bone marrow
12	352	5.6	68	22	AAW31616	Peptide #5653 enco
13	352	5.6	68	23	ABG41140	Human peptide enco
14	264	4.2	545	19	AAW79296	Neisseria UDP-N-ac
15	171.5	2.7	2469	23	AAE18207	Human MOLA protei
16	171.5	2.7	10182	23	ABP38314	Staphylococcus epi
17	168.5	2.7	5024	22	AAW82935	S. epidermidis ope
18	166.5	2.6	2471	20	AAW06816	Human Notch2 (humN
19	163.5	2.6	3135	15	AAW57474	P. falciparum tran
20	163.5	2.6	3135	21	AAW18223	Plasmodium falcipa
21	162.5	2.6	1654	6	AAW50777	Sequence of the P1
22	162.5	2.6	1979	21	AAW18171	Plasmodium falcipa
23	160	2.5	1963	22	AAW79838	Human protein SEQ
24	157.5	2.5	1872	19	AAW68510	Partial human Notc
25	155.5	2.5	2321	19	AAW49698	Human Notch3 prote
26	154.5	2.4	1254	11	AAW07503	Merizite apical-en
27	154.5	2.4	1254	18	AAW24575	Merizite apical-e
28	153.5	2.4	1014	23	ABG61923	Prostate cancer-as
29	153.5	2.4	1960	22	AAW78854	Human protein SEQ
30	153.5	2.4	2143	22	ABG01716	Novel human diagno
31	153	2.4	1964	20	AAW95557	Mus musculus notch
32	152.5	2.4	7201	22	ABW71136	Drosophila melanog
33	150.5	2.4	1639	19	AAW54145	P. falciparum synt
34	150.5	2.4	2663	22	AAW39097	Human polypeptide
35	150.5	2.4	2688	22	AAW40883	Human polypeptide
36	148	2.3	2633	22	ABG06505	Novel human diagno
37	147.5	2.3	2954	20	AAW01632	Amino acid sequenc
38	147	2.3	3685	10	AAW90290	Human Duchenne mus
39	145.5	2.3	2444	23	ABW07821	Constitutively act
40	145	2.3	28	22	AAW68261	Amino terminal of
41	145	2.3	962	22	AAW84956	Shrimp white spot
42	143.5	2.3	2285	20	AAW98149	Bacillus subtilis
43	143	2.3	1078	13	AAW28963	Notch hN3k full le
44	143	2.3	6815	22	ABW66811	Drosophila melanog
45	142	2.2	1109	22	AAU33434	Enterococcus faeca

#### ALIGNMENTS

RESULT 1

ABB30279

ID ABB30279 standard; Peptide; 367 AA.

XX ABB30279;

AC ABB30279;

XX 01-FEB-2002 (first entry)

DT Peptide #2930 encoded by breast cell single exon nucleic acid probe.

DE Human; microarray; single exon probe; gene expression; breast;

XX Human; microarray; single exon probe; gene expression; breast;

KW disease; cancer.

XX Homo sapiens.

OS WO200157271-A2.

PN 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000662.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000US-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PA Penn SG, Hanzel DK, Chen W, Rank DR;



DR WPI; 2001-496933/54.

XX New spatially-addressable set of single exon nucleic acid probes,  
PT useful for measuring gene expression in sample derived from human  
PT breast, comprises number of single exon nucleic acid probes -

XX Claim 27; SEQ ID NO 13247; 327pp + sequence listing; English.

XX The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived  
CC from human breast and BT 474 cells. The method involves contacting  
CC the probes with a collection of detectably labelled nucleic acids  
CC derived from mRNA of human breast, and then measuring the label  
CC bound to each probe of the microarray. The probes are useful for  
CC verifying the expression of regions of genomic DNA predicted to  
CC encode proteins. They are useful for gene discovery, and for  
CC determining predisposition and/or prognosing breast disease. Gene  
CC expression analysis is useful for assessing the toxicity of chemical  
CC agents on cells. The microarray of this invention presents a far greater  
CC diversity of probes for measuring gene expression, with far less bias  
CC than expressed sequence tag microarrays. The method is suitable for  
CC rapid production of functional information from genomic sequence. The  
CC present sequence is a peptide encoded by a single exon nucleic acid  
CC probe of the invention.

CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 367 AA;

Query Match 29.9%; Score 1893; DB 22; Length 367;

Best Local Similarity 99.7%; Pred. No. 4.1e-140;

Matches 366; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 529 HFHELYKVILLPNQTHYIIPKGECLPYFSAEVAKRGVGYSDNPIIRHASTANKWKTI 588

Db 1 HFHELYKVILLPNQTHYIIPKGECLPYFSAEVAKRGVGYSDNPIIRHASTANKWKTI 60

QY 589 HLIHSGMNATTIHFNLTFQNTNDEEFKMOITVEVDTRGPKLNSTAKGYENLVSPITL 648

Db 61 HLIHSGMNATTIHFNLTFQNTNDEEFKMOITVEVDTRGPKLNSTAKGYENLVSPITL 120

QY 649 LPEAEILFEDIPKEKRFKRDVNSTRAQEEVKIPLVNLISLLPKDAQLSLNTLDLQL 708

Db 121 LPEAEILFEDIPKEKRFKRDVNSTRAQEEVKIPLVNLISLLPKDAQLSLNTLDLQL 180

QY 709 EHGDIITLKGYNLSKALLRSFLMNSQHAKIKKQAIITDETNDLSLVAQEKQVHKSILPNS 768

Db 181 EHGDIITLKGYNLSKALLRSFLMNSQHAKIKKQAIITDETNDLSLVAQEKQVHKSILPNS 240

QY 769 LGVSRERLQRTTTPAVSVKNGHDQGNPDLDTETARFRVETHQTKIGNVTKEKPPSL 828

Db 241 LGVSRERLQRTTTPAVSVKNGHDQGNPDLDTETARFRVETHQTKIGNVTKEKPPSL 300

QY 829 IVPLESQMTKEKKITGKEKENSMEENAEHIGVTEVLLGRKLQHYTDSYLGFLPWEKKK 888

Db 301 IVPLESQMTKEKKITGKEKENSMEENAEHIGVTEVLLGRKLQHYTDSYLGFLPWEKKK 360

QY 889 YFLDLLD 895

Db 361 YFQDLLD 367

RESULT 2

AAW56268

ID AAW56268 standard; Protein; 367 AA.

XX AAW56268;

AC AAW56268;

XX 05-NOV-2001 (first entry)

DT Human brain expressed single exon probe encoded protein SEQ ID NO: 28373.

XX

KW Human; brain expressed exon; gene expression analysis; probe;  
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
XX epilepsy; cancer.  
OS Homo sapiens.  
XX  
PN WO200157275-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00667.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483446/52.

XX Single exon nucleic acid probes for analyzing gene expression in human

XX brains -

XX Example 4; SEQ ID NO: 28373; 650pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid

XX probes which are derived from genomic sequences expressed in the human

XX brain. They can be used to measure gene expression in brain cell samples,

XX which may enable the diagnosis and improved treatment of nervous system

XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,

XX epilepsy and cancers. The present sequence is a protein encoded by one of

XX the probes of the invention.

XX Sequence 367 AA;

Query Match 29.9%; Score 1893; DB 22; Length 367;

Best Local Similarity 99.7%; Pred. No. 4.1e-140;

Matches 366; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 529 HFHELYKVILLPNQTHYIIPKGECLPYFSAEVAKRGVGYSDNPIIRHASTANKWKTI 588

Db 1 HFHELYKVILLPNQTHYIIPKGECLPYFSAEVAKRGVGYSDNPIIRHASTANKWKTI 60

QY 589 HLIHSGMNATTIHFNLTFQNTNDEEFKMOITVEVDTRGPKLNSTAKGYENLVSPITL 648

Db 61 HLIHSGMNATTIHFNLTFQNTNDEEFKMOITVEVDTRGPKLNSTAKGYENLVSPITL 120

QY 649 LPEAEILFEDIPKEKRFKRDVNSTRAQEEVKIPLVNLISLLPKDAQLSLNTLDLQL 708

Db 121 LPEAEILFEDIPKEKRFKRDVNSTRAQEEVKIPLVNLISLLPKDAQLSLNTLDLQL 180

QY 709 EHGDIITLKGYNLSKALLRSFLMNSQHAKIKKQAIITDETNDLSLVAQEKQVHKSILPNS 768

Db 181 EHGDIITLKGYNLSKALLRSFLMNSQHAKIKKQAIITDETNDLSLVAQEKQVHKSILPNS 240

QY 769 LGVSRERLQRTTTPAVSVKNGHDQGNPDLDTETARFRVETHQTKIGNVTKEKPPSL 828

Db 241 LGVSRERLQRTTTPAVSVKNGHDQGNPDLDTETARFRVETHQTKIGNVTKEKPPSL 300

QY 829 IVPLESQMTKEKKITGKEKENSMEENAEHIGVTEVLLGRKLQHYTDSYLGFLPWEKKK 888

Db 301 IVPLESQMTKEKKITGKEKENSMEENAEHIGVTEVLLGRKLQHYTDSYLGFLPWEKKK 360

QY 889 YFLDLLD 895

Db 361 YFQDLLD 367

```
RESULT 3
AA016457
ID  AA016457 standard; Protein; 367 AA.
XX
AC  AA016457;
XX
DT  12-OCT-2001 (first entry)
XX
DE  Peptide #2891 encoded by probe for measuring cervical gene expression.
XX
KW  Probe; human; microarray; gene expression; cervical epithelial cell;
KW  cervical cancer.
XX
OS  Homo sapiens.
XX
PN  WO200157278-A2.
XX
PD  09-AUG-2001.
XX
PF  30-JAN-2001; 2001WO-US006070.
XX
PR  04-FEB-2000; 2000US-0180312.
PR  26-MAY-2000; 2000US-0207456.
PR  30-JUN-2000; 2000US-0608408.
PR  03-AUG-2000; 2000US-0632366.
PR  21-SEP-2000; 2000US-0234687.
PR  27-SEP-2000; 2000US-0236359.
PR  04-OCT-2000; 2000GB-0024263.
XX
PA  (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI  Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR  WPI; 2001-488901/53.
XX
PT  Human genome-derived single exon nucleic acid probes useful for
PT  analyzing gene expression in human cervical epithelial cells -
XX
PS  Claim 27; SEQ ID No 21283; 487pp; English.
XX
CC  The present invention relates to human single exon nucleic acid probes
CC  (SENP: see AA010068-AA128459). The present sequence is a peptide encoded
CC  by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CC  can be used to produce a single exon microarray, which can be used for
CC  measuring human gene expression in a sample derived from human cervical
CC  epithelial cells. By measuring gene expression, the probes are therefore
CC  useful in grading and/or staging of diseases of the cervix, notably
CC  cervical cancer.
CC
CC  Note: The sequence data for this patent did not form part of the printed
CC  specification, but was obtained in electronic format directly from WIPO
CC  at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ  Sequence 367 AA;

Query Match 29.9%; Score 1893; DB 22; Length 367;
Best Local Similarity 99.7%; Pred. No. 4.1e-140;
Matches 366; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 529 HFHELYKVILLPNQTHYIIPKGECLPYFSFAEVAKRGVEGAYSDNPIIRHASIANKWKTI 588
DB 1 HFHELYKVILLPNQTHYIIPKGECLPYFSFAEVAKRGVEGAYSDNPIIRHASIANKWKTI 60
QY 589 HLIHSGMGNATTIHFNLTFQNTNDEEFKMQITVEVDTRGPKLNSTAQKGYENLVSPITL 648
DB 61 HLIHSGMGNATTIHFNLTFQNTNDEEFKMQITVEVDTRGPKLNSTAQKGYENLVSPITL 120
QY 649 LPAEILFEDIPKEKRPFKRHVDNSTRRAQEEVKIPLVNTISLLPKDAQLSLNTLDLQL 708
DB 121 LPAEILFEDIPKEKRPFKRHVDNSTRRAQEEVKIPLVNTISLLPKDAQLSLNTLDLQL 180
QY 709 EHGDTILKGYNLKSALLSFLMNSQHAKIKNQAIITDETNDLSLVAPOEKQVHKSLPNS 768
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 181 EHGDTILKGYNLKSALLSFLMNSQHAKIKNQAIITDETNDLSLVAPOEKQVHKSLPNS 240
QY 769 LGVSERLQRLTFPAVSVKVNGHDQGNPDLDETTARFRVETHQTKTIGGNVTKEKPPSL 828
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 LGVSERLQRLTFPAVSVKVNGHDQGNPDLDETTARFRVETHQTKTIGGNVTKEKPPSL 300
QY 829 IVPLESQMTKEKKITGKEKENSMEENHIGVTEVLLGRKLOHYTDSYLGFLPWEKKK 888
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 IVPLESQMTKEKKITGKEKENSMEENHIGVTEVLLGRKLOHYTDSYLGFLPWEKKK 360
QY 889 YFLDLLD 895
      || |||||
Db 361 YFQDLLD 367

RESULT 4
AA028952
ID  AA028952 standard; Protein; 367 AA.
XX
AC  AA028952;
XX
DT  17-OCT-2001 (first entry)
XX
DE  Peptide #2989 encoded by probe for measuring placental gene expression.
XX
KW  Probe; microarray; human; placenta; antenatal diagnosis;
KW  genetic disorder.
XX
OS  Homo sapiens.
XX
PN  WO200157272-A2.
XX
PD  09-AUG-2001.
XX
PF  30-JAN-2001; 2001WO-US00663.
XX
PR  04-FEB-2000; 2000US-0180312.
PR  26-MAY-2000; 2000US-0207456.
PR  30-JUN-2000; 2000US-0608408.
PR  03-AUG-2000; 2000US-0632366.
PR  21-SEP-2000; 2000US-0234687.
PR  27-SEP-2000; 2000US-0236359.
PR  04-OCT-2000; 2000GB-0024263.
XX
PA  (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI  Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR  WPI; 2001-488901/53.
XX
PT  Human genome-derived single exon nucleic acid probes useful for
PT  analyzing gene expression in human placenta -
XX
PS  Claim 27; SEQ ID No 29221; 654pp; English.
XX
CC  The present invention relates to single exon nucleic acid probes (SENP:
CC  see AA131315-AA157546). The present sequence is a peptide encoded by one
CC  such probe. The probes are useful for producing a microarray for
CC  predicting, measuring and displaying gene expression in samples derived
CC  from human placenta. The probes are useful for antenatal diagnosis of
CC  human genetic disorders.
XX
SQ  Sequence 367 AA;

Query Match 29.9%; Score 1893; DB 22; Length 367;
Best Local Similarity 99.7%; Pred. No. 4.1e-140;
Matches 366; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 529 HFHELYKVILLPNQTHYIIPKGECLPYFSFAEVAKRGVEGAYSDNPIIRHASIANKWKTI 588
DB 1 HFHELYKVILLPNQTHYIIPKGECLPYFSFAEVAKRGVEGAYSDNPIIRHASIANKWKTI 60
QY 589 HLIHSGMGNATTIHFNLTFQNTNDEEFKMQITVEVDTRGPKLNSTAQKGYENLVSPITL 648
```

|||||  
 61 HLHSGMNAATHIFNLTFONTNDEEFKMQITVEVDREGPKLNSTAKQGYENLVSPITL 120  
 649 LPEAEILFEDIPKEKRPKPKRHDVNSTRAQEVKIPLVNISLLPKDAQLSLNTLDLQ 708  
 121 LPEAEILFEDIPKEKRPKPKRHDVNSTRAQEVKIPLVNISLLPKDAQLSLNTLDLQ 180  
 709 EHGDTITLKGYNLSKALLRSFLMNSQHAQIKNQAIITDETNDLSLVAPOEKQVHKSILPNS 768  
 181 EHGDTITLKGYNLSKALLRSFLMNSQHAQIKNQAIITDETNDLSLVAPOEKQVHKSILPNS 240  
 769 LGVSRERLQRLTFPAVSVKVNGHDQGNPDLDTTARFRVETHQKTIGGNVTKEKPPSL 828  
 241 LGVSRERLQRLTFPAVSVKVNGHDQGNPDLDTTARFRVETHQKTIGGNVTKEKPPSL 300  
 829 IVPLESQMTKEKKITGKEKENSMEENAHGVTVEVLLGRKLQHYTDSYLGFLPWEKKK 888  
 301 IVPLESQMTKEKKITGKEKENSMEENAHGVTVEVLLGRKLQHYTDSYLGFLPWEKKK 360  
 889 YFLDLLD 895  
 361 YFQDLLD 367

## RESULT 5

AAM04186  
 ID AAM04186 standard; Protein; 367 AA.

XX AC AAM04186;

XX DT 09-OCT-2001 (first entry)

XX DE Peptide #2868 encoded by probe for measuring breast gene expression.

XX KW Probe; human; breast disease; proliferative breast disease; non-carcinoma tumour.

XX KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.

XX XX Homo sapiens.

XX XX WO200157270-A2.

XX XX 09-AUG-2001.

XX XX 29-JAN-2001; 2001WO-US00661.

XX XX 04-FEB-2000; 2000US-0180312.

XX XX 26-MAY-2000; 2000US-0207456.

XX XX 30-JUN-2000; 2000US-0608408.

XX XX 03-AUG-2000; 2000US-0632366.

XX XX 21-SEP-2000; 2000US-0234687.

XX XX 27-SEP-2000; 2000US-0236359.

XX XX 04-OCT-2000; 2000GB-0024263.

XX XX (MOLE-) MOLECULAR DYNAMICS INC.

XX XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX XX WPI; 2001-476286/51.

XX XX Novel single exon nucleic acid probe used to measuring gene expression

XX XX in a human breast -

XX XX Claim 27; SEQ ID No 12926; 322pp; English.

XX XX The present invention relates to novel single exon nucleic acid probes

XX XX (see AAT00010-AA110067). The present sequence is a peptide encoded by one

XX XX such probe. The probes are useful for measuring human gene expression in

XX XX a human breast sample, where the probe hybridises at high stringency to a

XX XX nucleic acid expressed in the human breast. The probes are useful for

XX XX predicting, diagnosing, grading, staging, monitoring and prognosing

XX XX diseases of the human breast, particularly those diseases with polygenic

XX XX aetiology. The diseases include: breast cancer, disorders of proliferative

XX XX inflammatory diseases of the breast, fibrocystic changes, proliferative

CC breast disease and non-carcinoma tumours.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 367 AA;

Query Match 29.9%; Score 1893; DB 22; Length 367;  
 Best Local Similarity 99.7%; Pred No. 4.1e-140;  
 Matches 366; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 529 HFHELYKVILLPNQTHYIIPKGECLPYFSFAEVAKRGVEGAYSDNPIIRHASIAKWKTI 588  
 DB 1 HFHELYKVILLPNQTHYIIPKGECLPYFSFAEVAKRGVEGAYSDNPIIRHASIAKWKTI 60  
 QY 589 HLIMHSGMNAATHIFNLTFONTNDEEFKMQITVEVDREGPKLNSTAKQGYENLVSPITL 648  
 DB 61 HLIMHSGMNAATHIFNLTFONTNDEEFKMQITVEVDREGPKLNSTAKQGYENLVSPITL 120  
 QY 649 LPEAEILFEDIPKEKRPKPKRHDVNSTRAQEVKIPLVNISLLPKDAQLSLNTLDLQ 708  
 DB 121 LPEAEILFEDIPKEKRPKPKRHDVNSTRAQEVKIPLVNISLLPKDAQLSLNTLDLQ 180  
 QY 709 EHGDTITLKGYNLSKALLRSFLMNSQHAQIKNQAIITDETNDLSLVAPOEKQVHKSILPNS 768  
 DB 181 EHGDTITLKGYNLSKALLRSFLMNSQHAQIKNQAIITDETNDLSLVAPOEKQVHKSILPNS 240  
 QY 769 LGVSRERLQRLTFPAVSVKVNGHDQGNPDLDTTARFRVETHQKTIGGNVTKEKPPSL 828  
 DB 241 LGVSRERLQRLTFPAVSVKVNGHDQGNPDLDTTARFRVETHQKTIGGNVTKEKPPSL 300  
 QY 829 IVPLESQMTKEKKITGKEKENSMEENAHGVTVEVLLGRKLQHYTDSYLGFLPWEKKK 888  
 DB 301 IVPLESQMTKEKKITGKEKENSMEENAHGVTVEVLLGRKLQHYTDSYLGFLPWEKKK 360  
 QY 889 YFLDLLD 895  
 DB 361 YFQDLLD 367

## RESULT 6

ABG38224

ID ABG38224 standard; Peptide; 367 AA.

XX AC ABG38224;

XX DT 19-AUG-2002 (first entry)

XX DE Human peptide encoded by genome-derived single exon probe SEQ ID 27889.

XX KW Human; single exon probe; asthma; lung cancer; COPD; ILD;  
 KW chronic obstructive pulmonary disease; interstitial lung disease;  
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;  
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
 KW primary ciliary dyskinesia; pulmonary hypertension;  
 KW hyaline membrane disease.

XX XX Homo sapiens.

XX XX WO200186003-A2.

XX XX 15-NOV-2001.

XX PD 30-JAN-2001; 2001WO-US00665.

XX PF 04-FEB-2000; 2000US-180312P.

XX PR 26-MAY-2000; 2000US-207456P.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-234687P.

27-SEP-2000; 2000US-236359P.  
04-OCT-2000; 2000GB-0024263.  
(MOLE-) MOLECULAR DYNAMICS INC.  
Penn SG, Hanzel DK, Chen W, Rank DR;  
WPI; 2002-114183/15.  
Spatially-addressable set of single exon nucleic acid probes, used to  
measure gene expression in human lung samples -  
Claim 27; SEQ ID No 27889; 634pp; English.  
The invention relates to a spatially-addressable set of single exon  
nucleic acid probes for measuring gene expression in a sample derived  
from human lung comprising single exon nucleic acid probes having one of  
12614 nucleic acid sequences mentioned in the specification, or their  
complements or the 12387 open reading frames derived from the 12614  
probes. Also included are a microarray comprising the novel set of  
probes; the novel set of probes which hybridize at high stringency to a  
nucleic acid expressed in the human lung; measuring gene expression in a  
sample derived from human lung, comprising (a) contacting the array with  
a collection of detectably labeled nucleic acids derived from human lung  
mRNA, and (b) measuring the label detectably bound to each probe of  
the array; identifying exons in a eukaryotic genome, comprising  
(a) algorithmically predicting at least one exon from genomic sequences  
of the eukaryote; and (b) detecting specific hybridisation of detectably  
labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
having a fragment identical to the predicted exon, the probe is included  
in the above mentioned microarray; assigning exons to a single gene,  
comprising (a) identifying exons from genomic sequence by the method  
above and (b) measuring the expression of each of the exons in several  
tissues and/or cell types using hybridisation to a single exon  
microarrays having a probe with the exon, where a common pattern of  
expression of the exons in the tissues and/or cell types indicates that  
the exons should be assigned to a single gene; a peptide comprising one  
of 12011 sequences, mentioned in the specification, or encoded by the  
probes/open reading frames (ORF). The probes are used for gene  
expression analysis, and for identifying exons in a gene, particularly  
using human lung derived mRNA and for the study of lung diseases  
such as asthma, lung cancer, chronic obstructive pulmonary disease  
(COPD), interstitial lung disease (ILD), familial idiopathic pulmonary  
fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,  
Niemann-pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary  
haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,  
pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic  
pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension  
and hyaline membrane disease. The present sequence is a peptide/protein  
encoded by a single exon probe of the invention.  
Note: The sequence data for this patent did not form part  
of the printed specification, but was obtained in electronic  
format directly from WIPO at  
ftp.wipo.int/pub/published\_pct\_sequences.  
Sequence 367 AA;  
Query Match 29.9%; Score 1893; DB 23; Length 367;  
Best Local Similarity 99.7%; Pred. No. 4.1e-140;  
Matches 366; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 529 HFHELYKVILLNTHYIIPKGECLPYFSFAEVARGVGAGVSDNPITRHASIAKWKTI 588  
DB 1 HFHELYKVILLNTHYIIPKGECLPYFSFAEVARGVGAGVSDNPITRHASIAKWKTI 60  
QY 589 HLIMHSGMNATTHFNLTQNTNDEEFKMQITVEVDTRGPKLNSTAGKYNLSPTIL 648  
DB 61 HLIMHSGMNATTHFNLTQNTNDEEFKMQITVEVDTRGPKLNSTAGKYNLSPTIL 120  
QY 649 LPEAEILFEDIPKRFKFKHDVNSTRRRAQEEVKIPLVNISLPPKDAQSLNTLDLQ 708  
DB 121 LPEAEILFEDIPKRFKFKHDVNSTRRRAQEEVKIPLVNISLPPKDAQSLNTLDLQ 180

709 EHGDTLKGYNLSKALLRSFLMNSOHAHAKIKNOALITDETNDLSVAPOEKQVHKSILPNS 768  
DB 181 EHGDTLKGYNLSKALLRSFLMNSOHAHAKIKNOALITDETNDLSVAPOEKQVHKSILPNS 240  
QY 769 LGVSRERQLRTPPAVSVKVNGHDQGNPPLDLETTARFRVETHQTKTIGGNVTKEKPPSL 828  
DB 241 LGVSRERQLRTPPAVSVKVNGHDQGNPPLDLETTARFRVETHQTKTIGGNVTKEKPPSL 300  
QY 829 IVPLESQMTKEKKTITGKEKENSMEENAHNIGVTEVLLGRKLQHYTDSYLGFLPWEKKK 888  
DB 301 IVPLESQMTKEKKTITGKEKENSMEENAHNIGVTEVLLGRKLQHYTDSYLGFLPWEKKK 360  
QY 889 YFLDLLD 895  
DB 361 YFQDLLD 367  
RESULT 7  
ABB59094  
ID ABB59094 standard; Protein; 652 AA.  
XX  
AC ABB59094;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 4074.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX WPI; 2001-656860/75.  
XX DR N-PSDB; ABL03197.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Disclosure; SEQ ID NO 4074; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 652 AA;  
Query Match 13.5%; Score 854.5; DB 22; Length 652;  
Best Local Similarity 24.4%; Pred. No. 6e-58;  
Matches 220; Conservative 107; Mismatches 201; Indels 373; Gaps 15;  
QY 305 DEDISARFEDNEELRYSLRSIERHAPWVRNIFVINGQIPSWNLNDNPRVTIVTHQDVF 364





PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2001-488900/53.  
 XX Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human bone marrow -  
 XX  
 PS Example 4; SEQ ID NO: 31643; 658pp + Sequence Listing; English.  
 XX The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC bone marrow. They can be used to measure gene expression in bone marrow  
 CC samples, which may enable the improved diagnosis and treatment of cancers  
 CC such as lymphoma, leukaemia and myeloma. The present sequence is a  
 CC protein encoded by one of the probes of the invention.  
 XX  
 SQ Sequence 68 AA;  
 Query Match 5.6%; Score 352; DB 22; Length 68;  
 Best Local Similarity 100.0%; Pred. No. 4.8e-20;  
 Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 113 EKQLECLLTHCIKVPMLVLDLPALPANITLKDLPSPFSHSASDIFNVAKPKNPSTNVSV 172  
 Db 1 EKQLECLLTHCIKVPMLVLDLPALPANITLKDLPSPFSHSASDIFNVAKPKNPSTNVSV 60  
 Qy 173 VVFDSTKD 180  
 Db 61 VVFDSTKD 68  
 RESULT 12  
 AAM31616  
 ID AAM31616 standard; Protein; 68 AA.  
 XX AAM31616;  
 AC AAM31616;  
 DT 17-OCT-2001 (first entry)  
 XX Peptide #5653 encoded by probe for measuring placental gene expression.  
 DE Probe; microarray; human; placenta; antenatal diagnosis;  
 KW genetic disorder.  
 KW Homo sapiens.  
 OS  
 XX WO200157272-A2.  
 PN 09-AUG-2001.  
 PD 30-JAN-2001; 2001WO-US000663.  
 PF 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2002-114183/15.  
 XX Spatially-addressable set of single exon nucleic acid probes, used to  
 PT measure gene expression in human lung samples -

DR WPI; 2001-48897/53.  
 XX Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human placenta -  
 XX  
 PS Claim 27; SEQ ID No 31885; 654pp; English.  
 XX The present invention relates to single exon nucleic acid probes (SENP;  
 CC see AAI31315-AAI57546). The present sequence is a peptide encoded by one  
 CC such probe. The probes are useful for producing a microarray for  
 CC predicting, measuring and displaying gene expression in samples derived  
 CC from human placenta. The probes are useful for antenatal diagnosis of  
 CC human genetic disorders.  
 XX  
 SQ Sequence 68 AA;  
 Query Match 5.6%; Score 352; DB 22; Length 68;  
 Best Local Similarity 100.0%; Pred. No. 4.8e-20;  
 Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 113 EKQLECLLTHCIKVPMLVLDLPALPANITLKDLPSPFSHSASDIFNVAKPKNPSTNVSV 172  
 Db 1 EKQLECLLTHCIKVPMLVLDLPALPANITLKDLPSPFSHSASDIFNVAKPKNPSTNVSV 60  
 Qy 173 VVFDSTKD 180  
 Db 61 VVFDSTKD 68  
 RESULT 13  
 ABG41140  
 ID ABG41140 standard; Peptide; 68 AA.  
 XX ABG41140;  
 AC ABG41140;  
 DT 19-AUG-2002 (first entry)  
 XX Human peptide encoded by genome-derived single exon probe SEQ ID 30805.  
 DE Human; single exon probe; asthma; lung cancer; COPD; ILD;  
 KW chronic obstructive pulmonary disease; interstitial lung disease;  
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;  
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
 KW primary ciliary dyskinesia; pulmonary hypertension;  
 KW hyaline membrane disease.  
 XX  
 OS Homo sapiens.  
 XX WO200186003-A2.  
 PN 15-NOV-2001.  
 PD 30-JAN-2001; 2001WO-US000665.  
 PF 04-FEB-2000; 2000US-180312P.  
 PR 26-MAY-2000; 2000US-207456P.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-234687P.  
 PR 27-SEP-2000; 2000US-236359P.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2002-114183/15.  
 XX Spatially-addressable set of single exon nucleic acid probes, used to  
 PT measure gene expression in human lung samples -



XX PS Claim 27; SEQ ID NO 30805; 634pp; English.

XX CC The invention relates to a spatially-addressable set of single exon

XX CC nucleic acid probes for measuring gene expression in a sample derived

XX CC from human lung comprising single exon nucleic acid probes having one of

XX CC 12614 nucleic acid sequences mentioned in the specification, or their

XX CC complements or the 12387 open reading frames derived from the 12614

XX CC probes. Also included are a microarray comprising the novel set of

XX CC nucleic acid expressed in the human lung; measuring gene expression to a

XX CC sample derived from human lung, comprising (a) contacting the array with

XX CC a collection of detectably labeled nucleic acids derived from human lung

XX CC mRNA, and (b) measuring the label detectably bound to each probe of

XX CC the array; identifying exons in a eukaryotic genome, comprising

XX CC (a) algorithmically predicting at least one exon from genomic sequences

XX CC of the eukaryote; and (b) detecting specific hybridisation of detectably

XX CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,

XX CC having a fragment identical to the predicted exon, the probe is included

XX CC in the above mentioned microarray; assigning exons to a single gene,

XX CC comprising (a) identifying exons from genomic sequence by the method

XX CC above and (b) measuring the expression of each of the exons in several

XX CC tissues and/or cell types using hybridisation to a single exon

XX CC microarrays having a probe with the exon, where a common pattern of

XX CC expression of the exons in the tissues and/or cell types indicates that

XX CC the exons should be assigned to a single gene; a peptide comprising one

XX CC of 12011 sequences, mentioned in the specification, or encoded by the

XX CC probes/open reading frames (ORF). The probes are used for gene

XX CC expression analysis, and for identifying exons in a gene, particularly

XX CC using human lung derived mRNA and for the study of lung diseases

XX CC such as asthma, lung cancer, chronic obstructive pulmonary disease

XX CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary

XX CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,

XX CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary

XX CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,

XX CC pulmonary alveolar proteinosis, Karagenen syndrome, fibrocystic

XX CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension

XX CC and hyaline membrane disease. The present sequence is a peptide/protein

XX CC encoded by a single exon probe of the invention.

XX CC Note: The sequence data for this patent did not form part

XX CC of the printed specification, but was obtained in electronic

XX CC format directly from WIPO at

XX CC ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 68 AA;

Query Match 5.6%; Score 352; DB 23; Length 68;

Best Local Similarity 100.0%; Pred. No. 4.8e-20;

Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 113 EKOLECLTHCIKVPMLVDPALPANITLKDLPSPHFSASDFNVAKPNPSTNSV 172

DB 1 EKOLECLTHCIKVPMLVDPALPANITLKDLPSPHFSASDFNVAKPNPSTNSV 60

QY 173 VVFDSTKD 180

DB 61 VVFDSTKD 68

RESULT 14

AAW79296

ID AAW79296 standard; Protein; 545 AA.

XX AC AAW79296;

XX DT 15-FEB-1999 (first entry)

XX DE Neisseria UDP-N-acetyl-mannosamine (alpha1-6) polymerase.

XX KW Serogroup marker; molecular typing; capsular instability; vaccine;

XX KW meningitis; UDP-N-acetyl-mannosamine (alpha1-6) polymerase.

XX OS Neisseria meningitidis serogroup A strain F8229.

XX PN WO9845312-A1.

XX PD 15-OCT-1998.

XX PF 09-APR-1998; 98WO-US06946.

XX PR 23-SEP-1997; 97US-0936107.

XX PR 09-APR-1997; 97US-0069885.

XX PA (UYEM-) UNIV EMORY.

XX PI Stephens DS, Swartley JS;

XX DR WPI; 1998-583189/49.

XX DR N-PSDB; AAV69110.

XX PT New DNA sequences from *Neisseria meningitidis* involved in serotype A

XX PT polysaccharide synthesis - and recombinant sequences encoding

XX PT capsule-switching mutation(s) in *N. meningitidis*, for producing

XX PT polysaccharides useful in vaccines

XX PS Disclosure; Page 55-59; 111pp; English.

XX CC This polypeptide is encoded by ORF2, one of 4 open reading frames

XX CC identified in the region between the ctra and gale genes of the

XX CC encapsulated serogroup A *Neisseria meningitidis* strain F8229.

XX CC The ORF2 protein is predicted to be a UDP-N-acetyl-mannosamine

XX CC (alpha1-6) polymerase involved in the biosynthesis of the serogroup

XX CC A capsule. The ORF1 and ORF3-ORF4 gene products (see AAW79295 and

XX CC AAW79297-98) are also thought to be involved in capsule biosynthesis or

XX CC modification. The invention provides a model in which meningococcal

XX CC capsular serogroups are determined by specific genetic biosynthesis

XX CC cassettes that insert between the ctra operon and gale. For

XX CC serogroup A, the cassettes determining specificity of serogroups

XX CC can recombine to switch the type of capsule and serogroup expressed.

XX CC Such information is critical to the design of improved group A and

XX CC other meningococcal vaccines and to the understanding of the

XX CC molecular basis of serogroup A pathogenesis. The invention

XX CC provides recombinant meningococcal strains, recombinant DNA

XX CC constructs and immunological preparations useful as diagnostic

XX CC probes for detection and diagnosis of meningococcal diseases,

XX CC screening for specific serogroups and broad-based immunisations

XX CC with multivalent capsular polysaccharide conjugate vaccines.

XX SQ Sequence 545 AA;

Query Match 4.2%; Score 264; DB 19; Length 545;

Best Local Similarity 27.8%; Pred. No. 1.3e-11;

Matches 90; Conservative 52; Mismatches 130; Indels 52; Gaps 13;

QY 135 LPANITLKDLPSPHFSASDFNVAKPNPSTNSVVFSTKDVEDAHSGLLKG-NS 193

DB 97 LPSNETLK--PALCILESHKEDFLN-----KFLLTISSENKIQYFNGQIRPKS 145

QY 194 RQTVWRCY-----LTTDKVEPGLV-----LMQDLAFSGFPPTFKETNOLKT 235

DB 146 VNEIWTDFLSIAHVDKMLSTDRTLSSISQFWFRLECKEDKDFI-----LFTANRYSR 200

QY 236 KLPENLSKVKLLQLYSEASVALLKLNPKDFQELNKQTKNNMTDGEKELTISPAYLLWD 295

DB 201 KWKH---SIKNOLFKREG-----IRNYSSEISLSEEDHNFIDIDLVFTWVNSEDKNQW 251

QY 296 LSAISOSKQD---EDISARFEDNEELRYSLSRSTERHAPVNRNIFVTNGQIFSWNLND 352

DB 252 -ELYKKYKDPFNSDATSTSRFLSRDELKFAIRSWEMSGSFIRKIFIVSNCAPAWLLNN 310

QY 353 PRVTIVTHQDVFRNLHSLPTFSPAIESHVRIEGLSQKFIYLMNDVWFQKDVWPDFFYS 412

DB 311 PKIQWVYHEETMPQ-SALPTFSSHAIEITSLHHIPGISNYFIYSNDFLLTKPLNKDNFF 369

QY 413 HSKGQKVYL-TWVPVN--CAEGCP 433

Db 370 SNGIAKLRLKLEAWGNVNGECTEGEP 393

RESULT 15

AAE18207

ID AAE18207 standard; Protein; 2469 AA.

XX AC AAE18207;

XX DT 07-MAY-2002 (first entry)

XX DE Human MOL1a protein.

XX Secreted molecule; MOL1a protein; MOLX; cardiomyopathy; atherosclerosis;  
 KW diabetes; chromosomal disorder; albinism; anaemia; psoriasis; scarring;  
 KW liver cirrhosis; neurodegeneration; osteoarthritis; organ rejection;  
 KW cerebral thrombosis; hypertension; systemic lupus erythematosus; asthma;  
 KW immune disease; ischaemia; immunodeficiency; rheumatoid arthritis; ulcer;  
 KW HIV; human immuno deficiency virus; hepatitis; haemostatic disease; pain;  
 KW haematopoietic; thrombolytic; thrombocytopaenia; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; muscular disease; stress;  
 KW ocular disease; growth disorder; depression; epilepsy; contraceptive;  
 KW vulnery; osteopathic; haemostatic; tranquiliser; antidepressant;  
 KW analgesic; vasotropic; hypotensive; gene therapy.

XX OS Homo sapiens.

XX Key Location/Qualifiers

XX Peptide 1..25

XX Protein /label= Signal\_peptide

XX /label= Mature\_MOL1a\_protein

XX W0200206339-A2.

XX PD 24-JAN-2002.

XX PF 03-JUL-2001; 2001WO-US21249.

XX PR 03-JUL-2000; 2000US-215854P.

XX PR 03-JUL-2000; 2000US-215856P.

XX PR 03-JUL-2000; 2000US-215902P.

XX PR 07-JUL-2000; 2000US-216585P.

XX PR 07-JUL-2000; 2000US-216586P.

XX PR 07-JUL-2000; 2000US-216722P.

XX PR 17-JUL-2000; 2000US-218622P.

XX PR 17-JUL-2000; 2000US-218992P.

XX PR 27-JUL-2000; 2000US-221285P.

XX PR 14-FEB-2001; 2001US-268734P.

XX (CURA-) CURAGEN CORP.

XX Spaderna SK, Tchernev V, Liu X, Shenoy S, Spytek K, Zerhusen B;

XX Patturajan M, Taupier RJ, Rastelli L, Grosse WM, Szekeres ES;

XX Alsbrook J, Lepley DM, Shen L, Burgess CE, Shimkets RA;

XX Padigar M;

XX WPI; 2002-155038/20.

XX N-PSDB; AAD28941.

XX Nucleic acids encoding secreted polypeptides, designated MOLX

XX polypeptides, useful for treating a MOLX-associated disorder, e.g.

XX cardiomyopathy, atherosclerosis, diabetes and metabolic disorders -

XX Claim 1; Page 11; 223pp; English.

XX The patent discloses nucleic acid sequences encoding novel secreted

XX molecule (MOL) polypeptides, designated MOLX polypeptides (i.e. a MOL

XX protein where X is an integer from 1 to 8). Sequences of the invention

XX are useful for treating or preventing a MOLX-associated disorder in

XX humans. They are useful for treating or preventing cardiomyopathy,

XX atherosclerosis and disorders related to cell signal processing and

XX metabolic pathway modulation. The MOLX antibodies are useful for

CC treating or preventing diabetes and disorders related to cell signal  
 CC processing and metabolic pathway modulation. MOLX sequences are useful  
 CC for the treatment or diagnosis of other MOLX-associated disorders, e.g.  
 CC chromosomal disorders, albinism, anaemia, liver cirrhosis, psoriasis,  
 CC scarring, neurodegeneration, osteoarthritis, organ rejection, cerebral  
 CC thrombosis, ischaemia, hypertension, systemic lupus erythematosus, immune  
 CC diseases, immunodeficiency, HIV (human immuno deficiency virus), viral,  
 CC bacterial and fungal infections, hepatitis, rheumatoid arthritis, asthma,  
 CC haematopoietic, thrombolytic, haemostatic diseases, thrombocytopaenia,  
 CC ulcers, Alzheimer's disease, Parkinson's disease, Huntington's disease,  
 CC ocular disease, muscular diseases, growth disorders, loss of libido,  
 CC stress, depression, pain and epilepsy. They are useful for preventing  
 CC chemotherapy side effects and as contraceptives. Sequences of the  
 CC invention are also useful for gene therapy. The present sequence  
 CC is human Notch-like protein, MOL1a.

XX Sequence 2469 AA;

Query Match 2.7%; Score 171.5; DB 23; Length 2469;  
 Best Local Similarity 19.9%; Pred. No. 0.0033;  
 Matches 111; Conservative 63; Mismatches 176; Indels 207; Gaps 26;

QY 424 PVPNCAEGCPGSMWIKGYCDKACNNSACDWDGDCSGNNGSRYTAGGGTSGIGVGPW 483

Db 1419 PPATCLSQYCADKARDGVCDACNSHACQWDGDC-----SUTMENPW 1461

QY 484 -QFGGINSVSYCNQGC-----ANS-----WLADKF-----CDQACN 514

Db 1462 ANCSPLPCWDYINNQCDELNTVECLDFNFCQGNSTCKRYKACADHFNHCDQGCN 1521

QY 515 VLSCGFADGCGDHFHELYK-----VILLPNQTHIIPKGECLPYFSFAEV-----561

Db 1522 SEECGWDGLCAADQAPENLAEGTLVIVLMP-----PEQLQDARSFLRALTLLHTN 1574

QY 562 --AKRGEVAGSDNPILRIHASTANKWKTIIHIMHSGMNATTIHFNLTFONTNDEEFK---616

Db 1575 LRIKRDSQGLMVPYYPYGEKSAAMKKQ-----RMTRRSPLPGQEVEVA 1617

QY 617 -MQITVEVDTREGPKLNSTAQKGYENLYSPITLLPEAEILFEDIPKRRFPKFRHDVNS 675

Db 1618 GSKVFEIDNQK---CVQDSHCFKNTDAAALLAS-----HAIQG 1655

QY 676 TRRAQEEVKIPLVNT---SLLPKDAQLSLNTLDLQL-----EHGDIIT 714

Db 1656 T-----LSYPLVSVVSESLTPERTQL-LYLLAVAVIILFIILLGYIMAKRRKHGSLW 1708

QY 715 L-KGYNLSKALLRSFLMNSQAK-----IKNOAITDETNSLSVAPQEKQVHKS 763

Db 1709 LPEGFTLRDA-----SNHKRRPEPVGDVAVGLKNLSVQVSEAN--LI-----1748

QY 764 ILPNSIGVSRERLQRLTFPAVSVKVNHGDOG-----QNPLDLLETTARFRVETHHTOKTIGN 819

Db 1749 ---GTGTSEHWVDEGFPQK-KYKADEALLSEDDPIDR-----RPWTOQHLEAA 1795

QY 820 VTKKPPSLIVPLESQMTKE-----KKITGKEKENSMEENAEH---H 859

Db 1796 DIRTPPSLALTPQAEQEVLDVNVVRGPDCTPLMLASLRGSSDLSDDEDAEDSSAN 1855

QY 860 IGVTVELLGRKLIQHYTD 876

Db 1856 IITDLIVQGASLQAQTD 1872

Search completed: April 13, 2003, 03:20:30  
 Job time : 75.7772 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2003, 03:00:01 ; Search time 63.7872 Seconds  
(without alignments)  
3873.042 Million cell updates/sec

Title: US-10-023-888-2  
Perfect score: 6340  
Sequence: 1 METDTLLVLLWVPGSTG.....NRFLLMHELQEWRAIRDKLK 1199

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL\_21.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_virus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4248	67.0	847	4 Q96N13	Q96n13 homo sapien
2	3127	49.3	663	4 Q9UL12	Q9ul12 homo sapien
3	1281.5	20.2	384	11 Q61340	Q61340 mus musculus
4	1046	16.5	248	4 Q9NPW9	Q9npw9 homo sapien
5	854.5	13.5	652	5 Q9V553	Q9v553 drosophila
6	854.5	13.5	666	5 Q8SX14	Q8sx14 drosophila
7	460	7.3	132	4 Q9NV34	Q9nv34 homo sapien
8	327	5.2	602	16 Q9L112	Q9l112 streptomyce
9	324.5	5.1	942	16 Q9L114	Q9l114 streptomyce
10	321.5	5.1	586	16 Q69853	Q69853 streptomyce
11	313	4.9	541	16 Q69851	Q69851 streptomyce
12	286	4.5	238	2 Q9EVX1	Q9evx1 streptococ
13	265	4.2	545	16 Q9JW18	Q9jw18 neisseria m
14	264	4.2	545	2 Q68215	Q68215 neisseria m
15	259.5	4.1	442	5 Q8SSQ9	Q8ssq9 dictyosteli
16	255	4.0	532	16 Q06628	Q06628 mycobacteri

17	253.5	4.0	542	16 Q50025	O50025 mycobacteri
18	229.5	3.6	366	2 Q9RGR0	Q9rgr0 neisseria m
19	227.5	3.6	373	2 Q51151	Q51151 neisseria m
20	190	3.0	1476	13 Q90285	Q90285 carassius a
21	184	2.9	2752	5 Q9BJY0	Q9bjy0 plasmodium
22	183	2.9	2867	5 Q9N2M3	Q9n2m3 plasmodium
23	182	2.9	638	13 Q42372	Q42372 brachydanio
24	180.5	2.8	2352	5 Q61240	Q61240 halocynthia
25	176	2.8	1422	5 Q00914	Q00914 plasmodium
26	176	2.8	1999	11 Q63731	Q63731 rattus norv
27	171	2.7	2524	5 Q9GPA5	Q9gpa5 brachioosto
28	169.5	2.7	762	13 Q42373	Q42373 brachydanio
29	167.5	2.6	523	5 Q9NE33	Q9ne33 leishmania
30	167.5	2.6	3144	5 Q9GTK4	Q9gtk4 plasmodium
31	167	2.6	1387	5 Q9GZ76	Q9gz76 plasmodium
32	166	2.6	1330	5 Q9BK46	Q9bk46 plasmodium
33	164	2.6	3254	5 Q9BK45	Q9bk45 plasmodium
34	163.5	2.6	763	16 Q8XJY4	Q8xjy4 clostridium
35	163.5	2.6	2771	5 Q26216	Q26216 plasmodium
36	162.5	2.6	1979	5 Q96133	Q96133 plasmodium
37	162	2.6	2434	8 Q9MTH5	Q9mth5 oenothera h
38	161.5	2.5	898	11 Q9Z220	Q9z220 rattus norv
39	161.5	2.5	1694	5 Q9T2T5	Q9t2t5 plasmodium
40	160.5	2.5	1694	5 Q9NHX1	Q9nhx1 plasmodium
41	160.5	2.5	1720	5 Q25922	Q25922 plasmodium
42	160.5	2.5	3259	4 Q14789	Q14789 homo sapien
43	159.5	2.5	1928	5 Q9U0H2	Q9u0h2 plasmodium
44	158.5	2.5	2104	16 Q8XM24	Q8xm24 clostridium
45	158	2.5	1194	13 Q9W737	Q9w737 gallus gall

## ALIGNMENTS

## RESULT 1

Q96N13	ID	Q96N13	PRELIMINARY;	PRT;	847 AA.
AC	Q96N13;				
DT	01-DEC-2001	(T-EMBLrel. 19, Created)			
DT	01-DEC-2001	(T-EMBLrel. 19, Last sequence update)			
DT	01-JUN-2002	(T-EMBLrel. 21, Last annotation update)			
DE	CNA FLJ31575 fis, clone NT2R12001846, moderately similar to basic domain/leucine zipper transcription factor (Fragment).				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H., Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.; "NEDO human cDNA sequencing project."				
RL	Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AK056137; BAB71102.1; -				
DR	InterPro: IPR000800; Notch.				
DR	Pfam: PF00066; notch; 1.				
FT	NON_TER	847	847		
SQ	SEQUENCE	847 AA;	95165 MW;	0A115015824733C5	CRG64;

Query Match	67.0%;	Score	4248;	DB	4;	Length	847;
Best Local Similarity	99.9%;	Pred. No.	1.1e-245;				
Matches	802;	Conservative	1;	Mismatches	0;	Indels	0;
QY	35	SRDQYHVLFDSDYRDNTAGKSFQNRCLCLPMPIDVYVYTWNGTDLLELKLQVREQMEEQ	94				
Db	45	SRDQYHVLFDSDYRDNTAGKSFQNRCLCLPMPIDVYVYTWNGTDLLELKLQVREQMEEQ	104				

```
QY 95 KAMEILGKNTTEPKKSEKQLECLLTHCIKVPMLVLDPALPANITLKDLPSPYFSHSA 154
DB 105 KAMEILGKNTTEPKKSEKQLECLLTHCIKVPMLVLDPALPANITLKDLPSPYFSHSA 164
QY 155 SDINFAKPNPSTNVSVVYFDSTKQVEDAHSGILLKGNRSQTVWRYGLTDTKEVPGVLVM 214
DB 165 SDINFAKPNPSTNVSVVYFDSTKQVEDAHSGILLKGNRSQTVWRYGLTDTKEVPGVLVM 224
QY 215 QDLAFLSGFPPTFKETNQLTKLPENLSSKVLLQLYSEASVALLKLNPKDFQELNKOT 274
DB 225 QDLAFLSGFPPTFKETNQLTKLPENLSSKVLLQLYSEASVALLKLNPKDFQELNKOT 284
QY 275 KKNMTIDGKELTISPAYLLWDLISAISQSKODEDISASRFEDNEELRSLRSIERHAPWVR 334
DB 285 KKNMTIDGKELTISPAYLLWDLISAISQSKODEDISASRFEDNEELRSLRSIERHAPWVR 344
QY 335 NIFIVTNGQIPSWNLNDNPRVTIVTHQDVFRNLNLSHLPFTSSPAIESHVHRIEGLSKFY 394
DB 345 NIFIVTNGQIPSWNLNDNPRVTIVTHQDVFRNLNLSHLPFTSSPAIESHVHRIEGLSKFY 404
QY 395 LNDVDMFGKDWDPDFYSHSGOKVYLTWPVPCAEGCPGSKWIKDGYCDKACNNSACDWD 454
DB 405 LNDVDMFGKDWDPDFYSHSGOKVYLTWPVPCAEGCPGSKWIKDGYCDKACNNSACDWD 464
QY 455 GDCSGNSGGSRYTAGGGTGSIGVGQPMQFGGGINSVSYCNQGCANSLWADKFCDOACN 514
DB 465 GDCSGNSGGSRYTAGGGTGSIGVGQPMQFGGGINSVSYCNQGCANSLWADKFCDOACN 524
QY 515 VLSGCFDAGCGQDHFHLYKVILLPNQTHYIIPKGECLPYFSAEVAKRGVEGAYSDNP 574
DB 525 VLSGCFDAGCGQDHFHLYKVILLPNQTHYIIPKGECLPYFSAEVAKRGVEGAYSDNP 584
QY 575 IIRASIANKWKTHLIMHSGMNTTIFNLTFQNTNDEEFKQITVEVDTRGPKLNST 634
DB 585 IIRASIANKWKTHLIMHSGMNTTIFNLTFQNTNDEEFKQITVEVDTRGPKLNST 644
QY 635 AOKGYENLVSPITLPAEILFEDIPKEKRPKFKRHDVNSTRAQEVKIPLVNISLSP 694
DB 645 AOKGYENLVSPITLPAEILFEDIPKEKRPKFKRHDVNSTRAQEVKIPLVNISLSP 704
QY 695 KDAQLSLNTLDLQLEHGDITLKGYNLSKALLRSFLMNSQHAKIKNOAIITDETNDLSVA 754
DB 705 KDAQLSLNTLDLQLEHGDITLKGYNLSKALLRSFLMNSQHAKIKNOAIITDETNDLSVA 764
QY 755 POEKQVHKSILPNSLGVSRQLRTPPAVSVKNGHDQGNPDLDTTARFRVETHQK 814
DB 765 POEKQVHKSILPNSLGVSRQLRTPPAVSVKNGHDQGNPDLDTTARFRVETHQK 824
QY 815 TIGGNVTKKPPSLIVPLESQMT 837
DB 825 TIGGNVTKKPPSLIVPLESQMT 847

RESULT 2
Q9ULL2 PRELIMINARY; PRT; 663 AA.
AC Q9ULL2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE K1AA1208 protein (Fragment).
GN K1AA1208.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=20039619; PubMed=10574462;
RT Nagase T., Ishikawa K., Kikuno R., Hirose M., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
```

```
RT for large proteins in vitro.*;
RL DNA Res. 6:337-345(1999).
DR EMBL: AB033034; BAA86522.1; -.
DR InterPro: IPR002048; EF-hand.
DR Pfam: PF00036; efhand; 1.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 663 AA; 76150 MW; 76B4E4B6242C2CA8 CRC64;

Query Match 49.3%; Score 3127; DB 4; Length 663;
Best Local Similarity 98.5%; Pred. No. 8.7e-179; Indels 0; Gaps 0;
Matches 603; Conservative 2; Mismatches 7;

QY 528 DHFHELYKVILLPNQTHYIIPKGECLPYFSAEVAKRGVEGAYSDNPRIIRHASTANKWT 587
DB 22 DHFHELYKVILLPNQTHYIIPKGECLPYFSAEVAKRGVEGAYSDNPRIIRHASTANKWT 81
QY 588 IHLIMHSGMNTTIFNLTFQNTNDEEFKQITVEVDTRGPKLNSTAKQYENLVSPIT 647
DB 82 IHLIMHSGMNTTIFNLTFQNTNDEEFKQITVEVDTRGPKLNSTAKQYENLVSPIT 141
QY 648 LPEAEILFEDIPKEKRPKFKRHDVNSTRAQEVKIPLVNISLPPKDAQLSLNTLDLQ 707
DB 142 LPEAEILFEDIPKEKRPKFKRHDVNSTRAQEVKIPLVNISLPPKDAQLSLNTLDLQ 201
QY 708 LEHGDITLKGYNLSKALLRSFLMNSQHAKIKNOAIITDETNDLSVAPOEKVHKSILPN 767
DB 202 LEHGDITLKGYNLSKALLRSFLMNSQHAKIKNOAIITDETNDLSVAPOEKVHKSILPN 261
QY 768 SLGVSRQLRTPPAVSVKNGHDQGNPDLDTTARFRVETHQKIGGNVTKKPPS 827
DB 262 SLGVSRQLRTPPAVSVKNGHDQGNPDLDTTARFRVETHQKIGGNVTKKPPS 321
QY 828 LIVPLESQMTREKKTITGKEKENSMEENAEHIGVTEVLLGRKLOHGYTDSYLGFLPWEKK 887
DB 322 LIVPLESQMTREKKTITGKEKENSMEENAEHIGVTEVLLGRKLOHGYTDSYLGFLPWEKK 381
QY 888 KYFDLLDEESLKTQLAYFTDTSKNRARKDTFADSLRYVYVKNILNSFGFTSRKVPAMH 947
DB 382 KYFDLLDEESLKTQLAYFTDTSKNRARKDTFADSLRYVYVKNILNSFGFTSRKVPAMH 441
QY 948 PHMDIRIVMOELQDMPEEFDKTSFKVHSDMDQAFSYFYILMSAVQPINISQVDFEV 1007
DB 442 PHMDIRIVMOELQDMPEEFDKTSFKVHSDMDQAFSYFYILMSAVQPINISQVDFEV 501
QY 1008 DTQDSGLVSDREIRTLATRIHELPLSLQDLTGLEHMLINCSKMLPADITQLNNIPPTQES 1067
DB 502 DTQDSGLVSDREIRTLATRIHELPLSLQDLTGLEHMLINCSKMLPADITQLNNIPPTQES 561
QY 1068 YYDPNLPVTKSLVTNCKPVTDKIHKAYKDKNRYRFEIMGEEETAFKMRITNVSHVQQL 1127
DB 562 YYDPNLPVTKSLVTNCKPVTDKIHKAYKDKNRYRFEIMGEEETAFKMRITNVSHVQQL 621
QY 1128 DDIRKNPKFVC 1139
DB 622 DDIRKNPKFVC 633

RESULT 3
Q61340 PRELIMINARY; PRT; 384 AA.
ID Q61340
AC Q61340;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Basic domain/leucine zipper transcription factor (Fragment).
GN MAFB.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
```

```

RX MEDLINE=95094266; PubMed=8001130;
RA Cordes S.P., Barsh G.S.;
RT "The mouse segmentation gene kr encodes a novel basic domain-leucine
RL zipper transcription factor.";
RL Cell 79:1025-1034(1994);
DR EMBL: L36434; AAA65688.1; -.
DR MGD: MGI:104555; Maf.
DR InterPro: IPR002048; EF-hand.
DR Pfam: PF00036; ehand; 1.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 384 AA; 43754 MW; 51F473C8807A7E55 CRC64;

Query Match
Best Local Similarity 67.8%; Score 1281.5; DB 11; Length 384;
Matches 271; Conservative 31; Mismatches 77; Indels 21; Gaps 6;

QY 634 TAQKGYENLVSPITLLPEAEILFEDIPKEKRPKRDHVNSTRRAQEVKIPLVNISLL 693
Db |||:|||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
3 TTQKAYESLVSPVTPLPQADVPFEDVPKEKRPKIRRHVDVNTATGRFQBEVKIPRVNISLL 62

QY 694 PKDAQSLNTLDLQLEHGDTITLKGYNLSKALLRSLFNLNSQHAKTKNOAITDETNSILV 753
Db |||:|||||:||||:||||:||||:||||:||||:||||:||||:||||:
63 PKEAQVRLNLDLQLEHGDTITLKGYNLSKALLRSLFNLNSQHAKTKNOAITDETNSILV 121

QY 754 APOEKQVHKSILPSLVSGVSRQLRTPFAPVYKVNHGQGNPPDLDTETARFRVETHTQ 813
Db |||:|||||:||||:||||:||||:||||:||||:||||:||||:||||:
122 VPQENPSHRR--PHGFAGEHRSERWTAETVTVKGRDHALNPPVPLETNARL-----AQ 174

QY 814 KITGNVYKEKPPSILVPLESOMTKEKIKTGKEKENSMEENAHIGVTEVLGRKIQH 873
Db |||:|||||:||||:||||:||||:||||:||||:||||:||||:||||:
175 PTIGTVSKENISPLVPPESHLP-----KEESDRAEGNA---VPVKELVPGRRCSR 224

QY 874 YTDVSLGFLPWEKKYFELDLDEESLKTOLAYFTDSKNRARKDFTADSLRYVKNILN 933
Db |||:|||||:||||:||||:||||:||||:||||:||||:||||:||||:
225 ITQA---FCPGKKKKYFOLLDAEESLKTOLAYFTDSKHTGRLQKDTFADSLRYVKNILN 281

QY 934 SKFGFTSRKVPAPMHMDIRVMOQLDMQFPEEFKTSFKHVRHSEDMQFAFSFYIYLS 993
Db |||:|||||:||||:||||:||||:||||:||||:||||:||||:||||:
282 SKFGFTSRKVPAPMHMDIRVMOQLDMQFPEEFKTSFKHVRHSEDMQFAFSFYIYLS 341

QY 994 AVQPLNISQVFDEVTDSGVLSDREITLATRIHELPLS 1033
Db |||:|||||:||||:||||:||||:||||:||||:||||:||||:||||:
342 AVQPLNISQVFDEVTDSGVLSDREITLATRIHDLPLT 381

RESULT 4
Q9NPW9 PRELIMINARY; PRT; 248 AA.
AC Q9NPW9
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE Hypothetical 29.3 kDa protein (Fragment).
GN DKFZF02B226.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MELANOMA;
RA Bloembergen H., Boecher M., Brandt P., Mewes H.W., Weil B., Wiemann S.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL359588; CAB94874.1; -.
DR InterPro: IPR002048; EF-hand.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 248 AA; 29336 MW; DAC4C4A6C4505522 CRC64;

Query Match
Best Local Similarity 16.5%; Score 1046; DB 4; Length 248;
Matches 100.0%; Pred. No. 5.8e-55;

```

```

Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 10;

QY 1003 VFDEVDTDQSGVLSDRIRIATRIHELPLSLQDLTGLEHMLNCSKMLPADITQLNNIP 1062
Db |||:|||||:||||:||||:||||:||||:||||:||||:||||:||||:
5 VFDEVDTDQSGVLSDRIRIATRIHELPLSLQDLTGLEHMLNCSKMLPADITQLNNIP 64

QY 1063 PTOESYDNPPLPVTKSLVTNCKPVTOKIKKAYKDKNKKYRFEIMGEEIEAFKIRTVNSH 1122
Db |||:|||||:||||:||||:||||:||||:||||:||||:||||:||||:
65 PTOESYDNPPLPVTKSLVTNCKPVTOKIKKAYKDKNKKYRFEIMGEEIEAFKIRTVNSH 124

QY 1123 VVGQLDDIRKNPRKFCVCLNDNIDHNHDKDAQTVKAVLRDFYESMFPPIPSQFELPREYRNR 1182
Db |||:|||||:||||:||||:||||:||||:||||:||||:||||:||||:
125 VVGQLDDIRKNPRKFCVCLNDNIDHNHDKDAQTVKAVLRDFYESMFPPIPSQFELPREYRNR 184

QY 1183 LHMHELOEWAYRDKLK 1199
Db |||:|||||:||||:||||:||||:||||:||||:||||:||||:||||:
185 LHMHELOEWAYRDKLK 201

RESULT 5
Q9V553 PRELIMINARY; PRT; 652 AA.
AC Q9V553
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE CG8027 protein.
GN CG8027.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RC MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananthan G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Bockstein P., Brothier P.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhao G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

```

RT "The genome sequence of Drosophila melanogaster.";

RL Science 287:2185-2195(2000).

DR EMBL: AE003834; AAF58967.1; -.

DR FlyBase: FBgn0033392; CG8027.

DR InterPro: IPR000800; Notch.

DR Pfam: PF00066; notch; 1.

DR SMART: SM00004; NL; 1.

SQ SEQUENCE 652 AA; 75752 MW; 2728764810039458 CRC64;

Query Match 13.5%; Score 854.5; DB 5; Length 652;

Best Local Similarity 24.4%; Pred. No. 6.7e-43;

Matches 220; Conservative 107; Mismatches 201; Indels 373; Gaps 15;

QY 305 DEDISASRFEDNEELRYSLRSIERHAPWVRNIFVINGQIPSWLNLDNPRVTIVTHQDVF 364

Db 72 DDKYDPSRFDKNEELRYSLRSLEKHAIRHVIYVINGQIPSWLDLSYERVTVPVHEVLA 131

QY 365 RNLSHLPTFSSPAIESHVHRIEGLSQFIYLNDDVDFGKDWDPDDYSHSKGQKQVLTWP 424

Db 132 PDPQLPTFSSSAIETFLHRIPKSLRFLYLNDDIFLGAPLYPEDLYTEAGVRVYQAW 191

QY 425 VPCNACGCGSWIKDGYCDKACNNSACDWDGDCSGNSGGSRYIAGGGTSGIGVGPWQ 484

Db 192 VPGCALDCPWTYIGDGACDRHCNIDACQFDGDCSETGPAS----- 232

QY 485 FGGGINSVSYCNOGCANSWLADKFCQACNVLSGDFDAGDGHFELYKVILLPNQTH 544

Db 233 -----DAH 235

QY 545 YIIPKGECLPYFSPAFAVAKRGVEGAYSDNPIIRHASIANKWKTIHLIMHSGMNATTIHN 604

Db 236 VIPSKEVLE----- 245

QY 605 LTFQNTNDEEFKMQITVEVDTRGPKLNSTAGKYENLVSPITLLPEAEILFEDIPKEKR 664

Db 246 -----VQP-AAVQSRV-----HR 258

QY 665 FPK-----FKRHDVNSTRAQAEVETPLVNIISLLPKDAQLSLNTLDLQLEHGDITLKY 718

Db 259 FPQMGLOKLFRRSSANF-----KD-----VNRHR 282

QY 719 NLSKALLRSLFMSQAHAKIKNOAITDETNDLSVAPOEKQVHKSILPNSLGSYERLQRL 778

Db 283 NVSTLKELRRIVERNFNKAALMS---LNPELETSSSEPOTTQRH----- 322

QY 779 TPAVSVKVNHDGQGNPPLDLETTARFRVETHQKTIGGNVTKEKPPSLIVPLESQMTK 838

Db 323 ----- 322

QY 839 EKKITGKEKENSMEENAEHNIGVTEVLLGRKLOHYTDSYLGFLPWEKKYFLDLDEEE 898

Db 323 -----GLRKE----- 327

QY 899 SLKTQLAYFTDSKNRARKDTPADSLRYVYNKILNSKFGFTSRKVPAPMHMIDRIVMQE 958

Db 328 -----DFKS-----STDIYSHSLIATNMLNRAIFGKARHVLAAHVLGFLDKDIVEA 373

QY 959 LQDMFPEEDKTSFKVHRSEDMQAFSYFYILMSAVQPLNISQVDFEVDQSGVLSDR 1018

Db 374 MORFHOQILDTAHQRFAPTDLOAYAFYYSFLMSETKVMSVEIFEFTDGSATWSDR 433

QY 1019 EIRTLATRIHELPSLODLTGLEHMLINCSK---MLPADITQLNNIPPTQESYDNP 1074

Db 434 EVRTFLTRIYPPPLDWSAMRYFEVQNCNTRNLGMHLKVDVVEHSTL--VYERYEDSNLP 491

QY 1075 PVTKSLVYNCKPVTDKTHKAYKDKNRYRFEIMGE--EIAFKMIRTNVSHVQQLDIRK 1132

Db 492 TITRDLVVRCPILIAEALAAAFARPKYNFHVSPKRTSHSNFMMLTSLNLTVEFSLORLR 551

QY 1133 NPKRFVCLNDNIDNH--KDAQTKAVLRDFVESMFPPSPQFELPREYRNFRLHMHLEQW 1191

Db 552 NPKRFVCLNDNIDNANGEDNEMVPHLEDFYLSFFPPRSKFEPLPPQYRNFESWRFQW 611

QY 1192 R 1192

Db 612 K 612

RESULT 6

Q8SX14 PRELIMINARY; PRT; 666 AA.

AC Q8SX14;

DT 01-JUN-2002 (TREMBLrel. 21, Created)

DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE RE35033p.

GN CG8027.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,

RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,

RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,

RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,

RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,

RA Celniker S.;

RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY089618; AAL50356.1; -.

SQ SEQUENCE 666 AA; 77745 MW; F6FDB6D1C39248 CRC64;

Query Match 13.5%; Score 854.5; DB 5; Length 666;

Best Local Similarity 24.4%; Pred. No. 6.9e-43;

Matches 220; Conservative 107; Mismatches 201; Indels 373; Gaps 15;

QY 305 DEDISASRFEDNEELRYSLRSIERHAPWVRNIFVINGQIPSWLNLDNPRVTIVTHQDVF 364

Db 86 DDKYDPSRFDKNEELRYSLRSLEKHAIRHVIYVINGQIPSWLDLSYERVTVPVHEVLA 145

QY 365 RNLSHLPTFSSPAIESHVHRIEGLSQFIYLNDDVDFGKDWDPDDYSHSKGQKQVLTWP 424

Db 146 PDPQLPTFSSSAIETFLHRIPKSLRFLYLNDDIFLGAPLYPEDLYTEAGVRVYQAW 205

QY 425 VPCNACGCGSWIKDGYCDKACNNSACDWDGDCSGNSGGSRYIAGGGTSGIGVGPWQ 484

Db 206 VPGCALDCPWTYIGDGACDRHCNIDACQFDGDCSETGPAS----- 246

QY 485 FGGGINSVSYCNOGCANSWLADKFCQACNVLSGDFDAGDGHFELYKVILLPNQTH 544

Db 247 -----DAH 249

QY 545 YIIPKGECLPYFSPAFAVAKRGVEGAYSDNPIIRHASIANKWKTIHLIMHSGMNATTIHN 604

Db 250 VIPSKEVLE----- 259

QY 605 LTFQNTNDEEFKMQITVEVDTRGPKLNSTAGKYENLVSPITLLPEAEILFEDIPKEKR 664

Db 260 -----VQP-AAVQSRV-----HR 272

QY 665 FPK-----FKRHDVNSTRAQAEVETPLVNIISLLPKDAQLSLNTLDLQLEHGDITLKY 718

Db 273 FPQMGLOKLFRRSSANF-----KD-----VNRHR 296

QY 719 NLSKALLRSLFMSQAHAKIKNOAITDETNDLSVAPOEKQVHKSILPNSLGSYERLQRL 778

Db 297 NVSTLKELRRIVERNFNKAALMS---LNPELETSSSEPOTTQRH----- 336

QY 779 TPAVSVKVNHDGQGNPPLDLETTARFRVETHQKTIGGNVTKEKPPSLIVPLESQMTK 838

Db 337 ----- 336

QY 839 EKKITGKEKENSMEENAEHNIGVTEVLLGRKLOHYTDSYLGFLPWEKKYFLDLDEEE 898



```

Db 337 -----GLRKE----- 341
QY 899 SLKTLAYTDSKNRKYRDTFADSLRVYVVKILSKGFTGSRKVPAPHHMIDRIVMOE 958
Db 342 -----DFKS-----STDIYSHSLIATNMLLNRAYGFKARHVLAHVGFLLDKDIVEA 387
QY 959 LODMPPEEDFKTSFHKVRHSEDMOFAFYFYLMSAVQPLNISQVDFEVDTDOSGVLSOR 1018
Db 388 MQRFRHQIILTAHORFRAPTDLOAFAYISFLMSETKVMSEVEIFDFDGDGSATWSDR 447
QY 1019 EIRTLATRIHELPSLQDLTGLEHMLNCSK----MLPADITQLNIPPTQESYDNP 1074
Db 448 EVRTFLTRIYQPPLDWSAMRYEEVVQCNTRNLGMHLKVDIVESHSTL--VYERYEDSNLP 505
QY 1075 PYTKSLVTNCKPVTDIKHKAYKDKKRYFEINGE--EEIAFKMIRTNVSHVVGQDDIRK 1132
Db 506 TITRDVLWRCPLLAEALANFAVRPKYNFHVSPKRTSHSNFMMLTSLNLTVEESLDRUR 565
QY 1133 NPKRFVCLNDNDH-NH-KDAQTVKAVLRDFEYSEMFPIPSQFELPREYRNRLHMHLEQEW 1191
Db 566 NPKRFNCINDNDANRGEDNEWMRHLEDFYLSFFPRKSKFELPPQYRNRFESWRDFORM 625
QY 1192 R 1192
Db 626 K 626

RESULT 7
Q9NV34 PRELIMINARY; PRT; 132 AA.
AC Q9NV34;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DE CDNA FLJ10959 fis, clone PLACE1000562.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Konai F., Hara K., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.;
RT "NEDO human cDNA sequencing project."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK001821; BAA91926.1; -.
SQ SEQUENCE 132 AA; 16083 MW; 92B5DEB4E062AD07 CRC64;

Query Match 7.3%; Score 460; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 2.6e-20;
Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1115 MIRTNVSHVVGOLDIRKNPKRFVCLNDNDHNDHDKAQTVKAVLRDFYESMFPPIPSQFEL 1174
Db 1 MIRTNVSHVVGOLDIRKNPKRFVCLNDNDHNDHDKAQTVKAVLRDFYESMFPPIPSQFEL 60
QY 1175 PREYRNRLHMHLEQEWYRDKLK 1199
Db 61 PREYRNRLHMHLEQEWYRDKLK 85

RESULT 8
Q9L112 PRELIMINARY; PRT; 602 AA.
AC Q9L112;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)

```

```

DE Hypothetical protein SCO2594.
GN SCO2594 OR SCC88.05C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Ouail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL139298; CAB75375.1; -.
KW Hypothetical protein.
SQ SEQUENCE 602 AA; 67501 MW; 720EAB86A31D943F CRC64;

Query Match 5.2%; Score 327; DB 16; Length 602;
Best Local Similarity 42.7%; Pred. No. 2.1e-11;
Matches 61; Conservative 33; Mismatches 45; Indels 4; Gaps 1;

QY 280 IDGKELTISPAYLLWDLISAISQSKQDEDISASRFEDNEELRYSLRSTERHAPVNRITFV 339
Db 297 VDGSD----PAWIRSAEFSDRPYHEAANAARYLSRDELRYSLRSLNYAPVWRVILV 352
QY 340 TNGQIPSLNDMPRTVITVTHQDVFRNLHLPTFSSPAIESHVHIEGLSKQFIYLNDDV 399
Db 353 TADQTPDLNTHDPRKLVKVSKEIFSEPTSLPTFNSHAIESQLHHDGLSEHFLYFNDDV 412
QY 400 MFKDVMPPDDFYSHSGQKQVILT 422
Db 413 MLGRETLPOHFFLPNGLGQYILS 435

RESULT 9
Q9L114 PRELIMINARY; PRT; 942 AA.
AC Q9L114;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Putative transferase.
GN SCO2592 OR SCC88.03C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Saunders D., Harris D.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";

```



```
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RA "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL: AL139298; CAB75373.1; -.
DR InterPro: IPR001296; Glycos_transf_1.
DR Pfam: PF00534; Glycos_transf_1; 1.
KW Transference.
SQ SEQUENCE 942 AA; 105063 MW; 108B737B98A97B80 CRC64;

Query Match 5.1%; Score 324.5; DB 16; Length 942;
Best Local Similarity 24.9%; Pred. No. 5.6e-11;
Matches 101; Conservative 75; Mismatches 156; Indels 73; Gaps 13;

QY 93 EOKAMREILGKNTPEPTKSEKQLECLLTHCIKVPMLVLPALPANITLKLPSLYPSFH 152
DB 428 DQAREVVIGASDRSLVRAGRLSE-----VRDLOGSEIVQRFETVVEALE 475

QY 153 SASDFINNAKPK--NPSTNVSVVFDSTKVEDAHSGLLKGNRQTVMRGYLTDTKEVPG 210
DB 476 SSGIPYVLLRRDNDPRRLAVDAAEQTR-VRKALAGAYEG--KAVYAECLKPRTHAPG 531

QY 211 LVLMQ-----DLAFLSGFPPTKFNQLK-----TKLPENLSK----- 244
DB 532 VLLAERLEAVGEVAGLRFRPVVSTTLRFGPAYGCDIEFWQVPEEGGQGFVAPLR 591

QY 245 -----VKLLQLYSEASVALLKLNPKDFQELNKQTKK-----NNTIDGKELTISPAYLLW 294
DB 592 PSAVGPKLPSTPDARTV---KDREYPTLEPLTRKLVSDIIFPDVAVTWVDDSDPRW 647

QY 295 D-----LSAISQKQDEISASRFEDNEELRYSLRSIERHAPWVRNIFVINGQIP 345
DB 648 QERRARRRAALGLEAESGDE--AARFRNRDELRYSLRSIAMFAPWIRKIYLVTDQTP 704

QY 346 SWNLNDPRVTIVTHODVFNRLSHLPFTSSPAIESHVHRIEGLSOKFIYLNDDVMEGKV 405
DB 705 EWLATEHEGIEVSHRDIFTDQDCLPFTNSHSIESQLHHIDGLSEQFLYLNDDVFIGRPV 764

QY 406 WPDIFYSHSKQKQVYLTPVPCNCAEGCPGSGWIKDGYCDKACNNSA 450
DB 765 GAQREFLENGASRFF--WSPTTVVPGEP-TEDEGIFYAAAKNNRA 806

RESULT 10
ID O69853 PRELIMINARY; PRT; 586 AA.
AC O69853:
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein SCO6023.
GN SCO6023 OR SC1C3.11.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RA "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL: AL139298; CAB75373.1; -.
DR InterPro: IPR001296; Glycos_transf_1.
DR Pfam: PF00534; Glycos_transf_1; 1.
KW Transference.
SQ SEQUENCE 942 AA; 105063 MW; 108B737B98A97B80 CRC64;

Query Match 5.1%; Score 321.5; DB 16; Length 586;
Best Local Similarity 46.3%; Pred. No. 4.3e-11;
Matches 62; Conservative 30; Mismatches 33; Indels 9; Gaps 3;

QY 280 IDGKELTISPAYLLW--DLAISQKQDEISASRFEDNEELRYSLRSIERHAPWVRNIF 337
DB 282 VDGND----PA---WRRRRSAYDGGYHAESANAARYISRDELRYSLRALEQNAQWVRHVH 334

QY 338 IVTNGOIPSWNLNDPRVTIVTHODVFNRLSHLPFTSSPAIESHVHRIEGLSOKFIYLNDD 397
DB 335 LVTDGQRPAWLNDSPRLTVVDHSEIFADPALPTFNHSAIESRLHKLHGLSEHFLYLNDD 394

QY 398 DVNFGKDVMPDDFY 411
DB 395 DMFLGRPTPDQDF 408

RESULT 11
ID O69851 PRELIMINARY; PRT; 541 AA.
AC O69851:
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein SCO6021.
GN SCO6021 OR SC1C3.09.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RA "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL: AL023702; CAA19233.1; -.
KW Hypothetical protein.
SQ SEQUENCE 541 AA; 60084 MW; 45E93981D16C6D2 CRC64;

Query Match 4.9%; Score 313; DB 16; Length 541;
Best Local Similarity 39.2%; Pred. No. 1.2e-10;
Matches 67; Conservative 31; Mismatches 63; Indels 10; Gaps 5;

QY 280 IDGKELTISPAYLLWDLAISQKQDEISASRFEDNEELRYSLRSIERHAPWVRNIF 339
DB 242 VDGND----PAKQKQAKGEVYHAESASDARFISRDELRYSLRHLFAPWIRNIVV 297

QY 340 TNGOIPSWNLNDPRVTIVTHODVFNRLSHLPFTSSPAIESHVHRIEGLSOKFIYLNDDV 399
DB 340 TNGOIPSWNLNDPRVTIVTHODVFNRLSHLPFTSSPAIESHVHRIEGLSOKFIYLNDDV 399
```

```
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RA "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL: AL023702; CAA19233.1; -.
KW Hypothetical protein.
SQ SEQUENCE 586 AA; 66009 MW; 75F98F0916484116 CRC64;

Query Match 5.1%; Score 321.5; DB 16; Length 586;
Best Local Similarity 46.3%; Pred. No. 4.3e-11;
Matches 62; Conservative 30; Mismatches 33; Indels 9; Gaps 3;

QY 280 IDGKELTISPAYLLW--DLAISQKQDEISASRFEDNEELRYSLRSIERHAPWVRNIF 337
DB 282 VDGND----PA---WRRRRSAYDGGYHAESANAARYISRDELRYSLRALEQNAQWVRHVH 334

QY 338 IVTNGOIPSWNLNDPRVTIVTHODVFNRLSHLPFTSSPAIESHVHRIEGLSOKFIYLNDD 397
DB 335 LVTDGQRPAWLNDSPRLTVVDHSEIFADPALPTFNHSAIESRLHKLHGLSEHFLYLNDD 394

QY 398 DVNFGKDVMPDDFY 411
DB 395 DMFLGRPTPDQDF 408

RESULT 11
ID O69851 PRELIMINARY; PRT; 541 AA.
AC O69851:
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein SCO6021.
GN SCO6021 OR SC1C3.09.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RA "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL: AL023702; CAA19233.1; -.
KW Hypothetical protein.
SQ SEQUENCE 541 AA; 60084 MW; 45E93981D16C6D2 CRC64;

Query Match 4.9%; Score 313; DB 16; Length 541;
Best Local Similarity 39.2%; Pred. No. 1.2e-10;
Matches 67; Conservative 31; Mismatches 63; Indels 10; Gaps 5;

QY 280 IDGKELTISPAYLLWDLAISQKQDEISASRFEDNEELRYSLRSIERHAPWVRNIF 339
DB 242 VDGND----PAKQKQAKGEVYHAESASDARFISRDELRYSLRHLFAPWIRNIVV 297

QY 340 TNGOIPSWNLNDPRVTIVTHODVFNRLSHLPFTSSPAIESHVHRIEGLSOKFIYLNDDV 399
DB 340 TNGOIPSWNLNDPRVTIVTHODVFNRLSHLPFTSSPAIESHVHRIEGLSOKFIYLNDDV 399
```

Db 298 TDDQVPAWNRDLPGARIATHREIFRPNEDLPTFNHSHSIESQLHHIEGLAEHFLYFNDDM 357  
QY 400 MFGKDWDPDDFVSHSKGQKVLTPVPN-CAEGCPGSGWIKDGYCDKACNNS 449  
Db 358 FMGRPVAPHSFFT-PNGTARY--FPSRNRIPQGAFAE--TDSVPDAACKNN 403

RESULT 12  
Q9EVX1 PRELIMINARY; PRT; 238 AA.  
AC Q9EVX1;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE Putative hexose transferase.  
GN CPSU.  
OS Streptococcus salivarius.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
OC Streptococcaceae; Streptococcus.  
OX NCBI\_TaxID=1304;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20519245; PubMed=11065358;  
RA Almiron-Roig E., Mulholland F., Gasson M.J., Griffin A.M.;  
RT "The complete cps gene cluster from Streptococcus thermophilus NCFB  
RT 2393 involved in the biosynthesis of a new exopolysaccharide.";  
RL Microbiology 146:2793-2802(2000).  
DR EMBL; Y17900; CAC18360.1; -  
KW Transferase.  
SQ SEQUENCE 238 AA; 29110 MW; A964250B01B78988 CRC64;

Query Match 4.5%; Score 286; DB 2; Length 238;  
Best Local Similarity 37.7%; Pred. No. 1.5e-09;  
Matches 58; Conservative 28; Mismatches 44; Indels 24; Gaps 3;

QY 262 NPKDFQELNKOYKNNMTIDGKELTISPAYLLWDSISQSKODEDISASRPEDNEELRY 321  
Db 11 NDPEFIREKNKYTPHNRKID-----NDED-NVHRYRDYGTFFY 47  
QY 322 SLRSIERHAPWNEIIVNGQIPSWLNDNPRVITVTHQDVFNLSHLPTSSPAIESH 381  
Db 48 WFRMVERHAPWNNIYLITNGQIPKWLNVNHPKLVKWRHEE-FIPKEYLPFIFNASEIMN 106  
QY 382 VHRIGLSOKFIYLDVDFGKDVDPDDFYSHSK 415  
Db 107 IHRIDGLSENFVLFNDMDYLIQDKYSDFFVNEK 140

RESULT 13  
Q9JW8 PRELIMINARY; PRT; 545 AA.  
AC Q9JW8;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE Putative capsule biosynthesis protein.  
GN SABC OR NMA0200.  
OS Neisseria meningitidis (serogroup A).  
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
OX NCBI\_TaxID=65699;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;  
MEDLINE=2022556; PubMed=10761919;  
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,  
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,  
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,  
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,  
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,  
RA Whitehead S., Spratt B.G., Barrall B.G.;  
RT "Complete DNA sequence of a serogroup A strain of Neisseria  
RT meningitidis 22491.";  
RL Nature 404:502-506(2000).

DR EMBL; AL162752; CAB83514.1; -  
KW Complete proteome.  
SQ SEQUENCE 545 AA; 64133 MW; 7D2564286C95CD3D CRC64;

Query Match 4.2%; Score 265; DB 16; Length 545;  
Best Local Similarity 27.8%; Pred. No. 9.2e-08;  
Matches 90; Conservative 53; Mismatches 129; Indels 52; Gaps 13;

QY 135 LPANITLKDLPSPYFHSASDIFNVAKPNPSTNVSVVYFSTKDVDAHSGLLKG-NS 193  
Db 97 LPSNLTGK--PALCILESHKEDFLN-----KFLLTISSENKLYQKFGQIKNPKS 145  
QY 194 ROTVWRGY-----LTTDKVEPGLV-----LMQDLAFLSGPPPTFKETNQLKT 235  
Db 146 VNEIWTDLFSIAHVDKMLSTDRTLSSISQFWFRLEFCKEDKDFI-----LFTANRYSR 200  
QY 236 KLPELSSKVKLLQLYSEASVALLKLNPKDFQELNKOYKNNMTIDGKELTISPAYLLWD 295  
Db 201 KLWKH---SIKNNQLFKEG-----IRNYSEISSLPYEDHNFIDILVFTWVNSEDKNWQ 251  
QY 296 LSASISQSKQD---EDISARFEDNEELRYSLRSTIERHAPWVRNIFIVTNGQIPSWLNDN 352  
Db 252 -ELYKKYKPDENSATSTSFSLRDELKFAIRSWEMNGSFIRKIFIVSNCAAPPAILDLNN 310  
QY 353 PRVTIVTHQDVFNLSHLPTSSPAIESHVRHIEGLSQFIYLDVDFGKDVDPDDFY 412  
Db 311 PKIQWVHEEIMPO- SALPTFSSHAIESLHHPISGINSYFIYSNDDFLTKPLKNDFFY 369  
QY 413 HSKGQKVL-TWVPN--CAEGCP 433  
Db 370 SNGIAKLRLEAWNGVNGECTEGEP 393

RESULT 14  
O68215 PRELIMINARY; PRT; 545 AA.  
ID O68215  
AC O68215;  
DT 01-AUG-1998 (TREMBlrel. 07, Created)  
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE SABC.  
GN SABC.  
OS Neisseria meningitidis (serogroup A).  
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
OX NCBI\_TaxID=65699;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=F8229;  
RX MEDLINE=98175678; PubMed=9515923;  
RA Swartley J.S., Liu L.-J., Miller Y.K., Martin L.E., Edupuganti S.,  
RA Stephens D.S.;  
RT "Characterization of the gene cassette required for biosynthesis of  
RT the (alpha1-5)-linked N-acetyl-D-mannosamine-1-phosphate capsule of  
RT serogroup A Neisseria meningitidis.";  
RL J. Bacteriol. 180:1533-1539(1998).  
DR EMBL; AF019760; AAC38286.1; -  
SQ SEQUENCE 545 AA; 64097 MW; 1BC3400C5CC33DCD CRC64;

Query Match 4.2%; Score 264; DB 2; Length 545;  
Best Local Similarity 27.8%; Pred. No. 1.1e-07;  
Matches 90; Conservative 52; Mismatches 130; Indels 52; Gaps 13;

QY 135 LPANITLKDLPSPYFHSASDIFNVAKPNPSTNVSVVYFSTKDVDAHSGLLKG-NS 193  
Db 97 LPSNLTGK--PALCILESHKEDFLN-----KFLLTISSENKLYQKFGQIKNPKS 145  
QY 194 ROTVWRGY-----LTTDKVEPGLV-----LMQDLAFLSGPPPTFKETNQLKT 235  
Db 146 VNEIWTDLFSIAHVDKMLSTDRTLSSISQFWFRLEFCKEDKDFI-----LFTANRYSR 200  
QY 236 KLPELSSKVKLLQLYSEASVALLKLNPKDFQELNKOYKNNMTIDGKELTISPAYLLWD 295  
Db 201 KLWKH---SIKNNQLFKEG-----IRNYSEISSLPYEDHNFIDILVFTWVNSEDKNWQ 251

1

Search completed: April 13, 2003, 03:22:39  
Job time : 69.7872 secs











QY 946 HMPHMDIRVMOELODMFPEEDKTSFKVHRSEDMQAFSYFYILMSAVQPLNISQVDF 1005  
Db 1274 -----LQKILIEENDMITEFEVKQLSEGVDSEIKQLPLSNLRYDFMEKREVNLIT---E 1325  
QY 1006 EVDTD-OSGVLSDREITRIATRIHELPLSLDGLTGLEHMLNCSKMLPADITQLNNIPT 1064  
Db 1326 EVKTEENENNEEL-AYDRENKILLWL-----LYOQGVVQIKNI-HI 1368  
QY 1065 QESYDPLNPVPTKSLVT-----NCKPVTDKIHKAYKDK 1098  
Db 1369 LTPYMPPEAYNFYKATTFENVEYRCRLIVDAIAELYAKK 1409

## RESULT 11

S21801

myosin heavy chain, neuronal [similarity] - rat

N:Alternate names: myosin II

N:Contains: myosin ATPase (EC 3.6.4.1)

C:Species: Rattus norvegicus (Norway rat)

C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 19-Apr-2002

C:Accession: S21801; PN0013; S18134

R:Sun, W.; Chantler, P.D.; 1992

J. Mol. Biol. 224, 1185-1193, 1992

A:Title: Cloning of the cDNA encoding a neuronal myosin heavy chain from mammalian brain

A:Reference number: S21801; MUID:92235856; PMID:1569576

A:Accession: S21801

A:Molecule type: mRNA

A:Residues: 1-1999 &lt;SUN&gt;

A:Cross-references: EMBL:X62659

R:Sun, W.; Chantler, P.D.

Biochem. Biophys. Res. Commun. 175, 244-249, 1991

A:Title: A unique cellular myosin II exhibiting differential expression in the cerebral

A:Reference number: PN0013; MUID:91151356; PMID:1998509

A:Accession: PN0013

A:Molecule type: mRNA

A:Residues: 1914-1998, 1' &lt;SU2&gt;

A:Experimental source: brain

C:Superfamily: myosin heavy chain; myosin motor domain homology

C:Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; nucleotide

F:84-763/Domain: myosin motor domain homology &lt;NMOT&gt;

F:174-181/Region: nucleotide-binding motif A (P-loop)

F:341-575/Region: actin binding #status predicted

F:653-675/Region: actin binding #status predicted

F:836-1999/Domain: coiled coil #status predicted &lt;COI&gt;

F:836-1276/Region: S2

F:1277-1999/Region: light meromyosin

F:125/Modified site: N6,N6-trimethyllysine (Lys) #status predicted

F:180/Binding site: ATP (Lys) #status predicted

F:693-703/Active site: Cys #status predicted

F:1916/Binding site: phosphate (Ser) (covalent) #status predicted

F:1943/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match

Best Local Similarity 2.8%; Score 176; DB 1; Length 1999;

Matches 136; Conservative 104; Mismatches 232; Indels 158; Gaps 28;

QY 612 DEEFKQIIVETDTRGPKLNSTAKGY-----ENLYSPITLLPEARILFEDIPK 661

Db 850 DEELMKK--VEKQTKVEALEMERKHQOLLEKNILAEQLQAEELFAEAEMEARLAA 907

QY 662 EKFPFKRHVDNSTRRAQEVKPIPLVNLISLLPKDAQLSLNLTLDLQLEHGDITLKGYNLS 721

Db 908 KQLELEILLDLSESVVEEERNQILQNEK---KKEQGHKNLDEQLDLMESARQKLQLE 964

QY 722 K----SALLR-----SFLMNSQAKI-KNOAITD---EFTNSLVAPOEQVHKSLPNS- 768

Db 965 KVTTEAKKLKEBEQIILEDQNKLAKEKKLLEDRIAETTNLTTEEESKSLAKLNKH 1024

QY 769 ----LGVSELQRLITFPAYSVKVNQDQONPLDLETTARFVETHQKTIGGNVTEK 824

Db 1025 EAMITDEELRR-----EKKQEQLE-----KTRRKLEGSDTSL 1060

QY 825 PPSLIVPLESOMTEKKKITKEKEN-----SRMEEN-AENHIGVTEVLLGRKLQHYTDSY 878  
Db 1061 --DOIAELOAQIAELKQOLAKKEELOALARVEEAAQNMALKKI---RELSQISEL 1115  
QY 879 LGFLPWK-----KKYFLDLLDEESLKTQLAYFTDS---KNRARYKRTDFADSLRYVN 929  
Db 1116 QEDLESERASRNKAOKRDLGEELEALKTELEDLTDTAAQQLRSKRE-----QEVN 1169  
QY 930 KILASKGFTSRKVPAMPHMIDR--IVMOELODMFPE----- 965  
Db 1170 -ILAKTLEEKATHEAQIQEMRQKSHQAVELEAQLEQKTKRKVANLEKAKQTLNENRGE 1228  
QY 966 -----EFDKTSFKVHRSEDMQAFSYFYILMSAVQPLNISQVDFDVTDOGSVLSD 1017  
Db 1229 ANEVKVLLOGGRDSEHKRKVE-----AQLEQLOVK--FNE-----GE 1264  
QY 1018 REIRTLATRIHELPLSLDGLTGLEHMLNCSKMLPADITOL--NNIPTQSYDNPPLPV 1076  
Db 1265 RVVTEADKVTQLQVELDNVTGLLSQSDSKSLTKDFSALESQLODTQELLOEENRQKL 1324  
QY 1077 TKSIVTNCKPVTDKIKAYDKNKYRFEIMGEERIAFKMIRTNVSHVVGOLDIDIRKNPRK 1136  
Db 1325 --SLSTKLKQVED-----EKNSFREQLEEEEEEKHNLEKQIATLHAQVADMKKMED 1375  
QY 1137 FV-CINDNIDHNHDKAQTVKAVLRDFFYEM 1165  
Db 1376 SVGCL-----ETAEEVKRKLOKLEGL 1397

## RESULT 12

A26868

major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C:Date: 19-Nov-1988 #sequence\_revision 19-Nov-1988 #text\_change 09-Jun-2000

C:Accession: A26868

R:Tanabe, K.; Mackay, M.; Goman, M.; Scalfe, J.G.

J. Mol. Biol. 195, 273-287, 1987

A:Title: Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmodium

A:Reference number: A26868; MUID:88011243; PMID:3079521

A:Accession: A26868

A:Molecule type: DNA

A:Residues: 1-1701 &lt;TAN&gt;

C:Superfamily: major merozoite surface antigen

C:Keywords: surface antigen

F:1-19/Domain: signal sequence #status predicted &lt;SIG&gt;

F:20-1701/Product: major merozoite surface antigen #status predicted &lt;MAT&gt;

Query Match

Best Local Similarity 2.7%; Score 171; DB 2; Length 1701;

Matches 246; Conservative 164; Mismatches 437; Indels 420; Gaps 62;

QY 107 EPTKKSEKQLECLIT--HCIKVPMVLDPALPANITLKDLPSPHSDIFENVAKPK 164

Db 27 ELVKLEALEDAVLVTGYSLFQKEKVMVNECTSGTAVTTSTPGSGSVTSGGSVASVA 86

QY 165 NPSNVSVVVFSDTKDVEDAHSGLLKGNROT-----VWRGYLTDD 205

Db 87 SGSGGVS-----ASGG--SGNSRRTNPSDSSNTKYATDLKRVQNYLFTI 133

QY 206 KEVPLGLVMQDLAFSGPPPTFKETNOLKTKLPENLSKVKLLQLYSEASVALLKLNPK 265

Db 134 KELK-----YPELFDLTNHLMT--LSKNVGVFKYLDIDYEEINELLYKLNFFY 179

QY 266 DF--QELNK-----QTKKNMTIDGRELITSPAYLWDLAISQSQD-----ED----- 307

Db 180 DLLRAKLMDACANSYQIIPFNKIRANELDVLKKIVFGYRKPLDNIKNDKYGMEDIKKN 239

QY 308 -----ISASRPEDNE-----LVSLSRISERHAPWRNFIPTVNGO 343

Db 240 KTTANINELLESGSKYIDQKNADNEEGKKLYQAQNLFIYKQLEAHNLISVLEKR 299

QY 344 IPSWLNLDNPRVTIVTHQDVFRNLSHLPTFSSP-----ATESHVHRIEGLSKQFIY 394

Db 300 IDTLKKNNIKKLEEDIDKIKTDAENPTTGSKPNPLPENKKKEVEGHEEKIEA-KTIK 358  
Qy 395 LNDVFMFGKDWPDYDFSHSGQKQVILT-----WVPNCAEGCPGSKIWDG 440  
Db 359 FNIDSLF-TDPLEYLYLREKNKVDVTPKSDPTKSVQIPKVPYPN-----404  
Qy 441 YCDKACNNSACDWDGDCSGSGSRYIAGGGTGSIGVGQPMQFGGINSVSYNOGCA 500  
Db 405 -----GIVYPLPLTDIHNLSAADND--K 425  
Qy 501 NSWLADKFCDOACNVLSGCFDAGD-CGQDHFHLYKYVILLPNQTHYIIPKGECLPYFSFA 559  
Db 426 NSY-----GDLMPDTEKEINKEITDNKERKI-----FI 455  
Qy 560 EVAKRGVEGAYSDNPIIRHASIANKWKTIHLHSGMNNATTI-----HFNLTQNTNDEE-- 614  
Db 456 NNKKOID---LEEKNNHTRKEQNK-KLEDEYKSKKQYEELEKFEYEMKNNFNDRKDV 511  
Qy 615 ---FKMQITVEVD-TREGPKLNSTAQGYENLVSPITLLPEAEILFEDIPKEKRPK--F 668  
Db 512 DKIFSARTYNNVEKORYNNKFPSS-----NNSVYVQVKKKALSYLEDSLRKGISEKDF 566  
Qy 669 KRHDVNST-----RRAQEEV-----KIPLVNISSLPKDAQLSLNTLD---LQLEHGDIT 714  
Db 567 NHYTLTKGLEADIKKLTTEEIKSENKILEKFNKGLTHSANASLEVS DIVKLOQKVLLI 626  
Qy 715 LKGYNLKSALLRSFLMNSQAHAKIKNOAILTD-----ETNDSLVAPOEKVHKSILP 766  
Db 627 KKIEDLRKIEL---FLKNAQ---LKDSIHVPNIYKPNKPEPYIYLIVLKEVDKLEKPEIP 680  
Qy 767 NSLGVSEY---LQRLTFPAVSVKVNHGQGNPPL-----DLETTARFRVETHTQKTIG 817  
Db 681 KYVDMKKEQAVLSITQPLVAASETTEDGGHSTHTLSQSGETEVEETEVE--TEETVG 737  
Qy 818 ---GNVTEKPPSLIVPLESQMTKEKKITGKEKENSMEENAHENHIGVTEVLLGRKLQHY- 874

## RESULT 13

A54498

major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (is

C:Species: plasmodium falciparum

C:Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 09-Jun-2000

C:Accession: A54498

R:Peterson, M.G.; Coppel, R.L.; McIntyre, P.; Langford, C.J.; Woodrow, G.; Brown, G.V  
Mol. Biochem. Parasitol. 27, 291-302, 1988  
A:Title: Variation in the precursor to the major merozoite surface antigens of Plasmo  
A:Reference number: A54498; MUID:88142999; PMID:2449612  
A:Accession: A54498  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1701 <PPT>  
A:Cross-references: GB:M19143; NID:g160412; PIDN:AAA29653.1; PID:g160413  
C:Superfamily: major merozoite surface antigen  
C:Keywords: surface antigen

Query Match 2.7%; Score 170; DB 2; Length 1701;

Best Local Similarity 19.4%; Pred. No. 0.27;

Matches 246; Conservative 164; Mismatches 437; Indels 420; Gaps 62;

Qy 107 EPTKSEKOLECLLT--HCIKVPMVLVDPALPANTITLDPLSPYSPFSHASDIFNWKPK 164

Db 27 ELVKLEALEDAVLGTYSLFQKEKMWLNEGTSGTAVTTSTPGSGSVTSGGVSASVASVA 86

Qy 165 NPSTNVVVVFDSTKDVEDAHSGLLKGNSROT-----VWRGYLTTD 205

Db 87 SGGSGSV-----ASGG--SGNSRRTPNSDSDSNKTYADLKHVQVNYLFTI 133

Qy 206 KEYPGLVLMQDLAFLSGFPPTKETNOLKTLPENLSSKVLLQLYSEASVALLKNNPK 265

Db 134 KELK-----YPELFDLTHHMLT-LSKNVDFGFKYLDGYYEINELLYKLNFY 179

Qy 266 DF--QELNK-----QTKKNTIDGKELTISPAYLLMDLSAISOSKOD-----ED----- 307

Db 180 DLLRAKLNDACANSYCOIPFNLKIRANELDVLKIVFGYRPLDNKIKONVKMEDYIKKN 239

Qy 308 -----ISASREFNEE-----LRYSLRSIERHAPWVRNIFVTNGO 343

Db 240 KTTIANINELIGSKKTDQNNADNEEGKKLYOAQYNLFYNNKQLOEHAHLNISVLEKR 299

Qy 344 IPSWLNLDNPRVTIYTHQDVFNLSHLPTFSSP-----AIESHVHREGLSOKFIY 394

Db 300 IDTLKKNNIKKLEEDIDKIKTDAENPTTGSKPNPLPENKKKEVEGHEEKIEA-KTIK 358

Qy 395 LNDVFMFGKDWPDYDFSHSGQKQVILT-----WVPNCAEGCPGSKIWDG 440

Db 359 FNIDSLF-TDPLEYLYLREKNKVDVTPKSDPTKSVQIPKVPYPN-----404

Qy 441 YCDKACNNSACDWDGDCSGSGSRYIAGGGTGSIGVGQPMQFGGINSVSYNOGCA 500

Db 405 -----GIVYPLPLTDIHNLSAADND--K 425

Qy 501 NSWLADKFCDOACNVLSGCFDAGD-CGQDHFHLYKYVILLPNQTHYIIPKGECLPYFSFA 559

Db 426 NSY-----GDLMPDTEKEINKEITDNKERKI-----FI 455

Qy 560 EVAKRGVEGAYSDNPIIRHASIANKWKTIHLHSGMNNATTI-----HFNLTQNTNDEE-- 614

Db 456 NNKKOID---LEEKNNHTRKEQNK-KLEDEYKSKKQYEELEKFEYEMKNNFNDRKDV 511

Qy 615 ---FKMQITVEVD-TREGPKLNSTAQGYENLVSPITLLPEAEILFEDIPKEKRPK--F 668

Db 512 DKIFSARTYNNVEKORYNNKFPSS-----NNSVYVQVKKKALSYLEDSLRKGISEKDF 566

Qy 669 KRHDVNST-----RRAQEEV-----KIPLVNISSLPKDAQLSLNTLD---LQLEHGDIT 714

Db 567 NHYTLTKGLEADIKKLTTEEIKSENKILEKFNKGLTHSANASLEVS DIVKLOQKVLLI 626

Qy 715 LKGYNLKSALLRSFLMNSQAHAKIKNOAILTD-----ETNDSLVAPOEKVHKSILP 766

Db 627 KKIEDLRKIEL---FLKNAQ---LKDSIHVPNIYKPNKPEPYIYLIVLKEVDKLEKPEIP 680

Qy 767 NSLGVSEY---LQRLTFPAVSVKVNHGQGNPPL-----DLETTARFRVETHTQKTIG 817

Db 681 KYVDMKKEQAVLSITQPLVAASETTEDGGHSTHTLSQSGETEVEETEVE--TEETVG 737

Qy 818 ---GNVTEKPPSLIVPLESQMTKEKKITGKEKENSMEENAHENHIGVTEVLLGRKLQHY- 874



Qy 782 AVSVKVNHGQGNPPPLDLETTARFVE---THTQKT---IGGNVTKEKPPSLIVPLES 834  
Db 1019 EPESPIKLHTEAAGSYAITTEITRESVNIIDPRHNRVTVLHWIASNSAEKSEDLIV---- 1074  
Qy 835 QMTREKKITGKEKENSMEEN 855  
Db 1075 HEAKECIAAGADVNMDCDEN 1095

Search completed: April 13, 2003, 03:23:40  
Job time : 49.8703 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2003, 01:36:25 ; Search time 22.9235 Seconds  
(without alignments)  
2169.392 Million cell updates/sec

Title: US-10-023-888-2

Perfect score: 6340

Sequence: 1 METDTLLVLLVLLVPGSTG.....NRFLHMHLEQEWRAVYDKLK 1199

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	171	2.7	1701	1 MSP1_PLAFM	P08569 plasmodium
2	170	2.7	1701	1 MSP1_PLAFM	P13819 plasmodium
3	169.5	2.7	2470	1 NTC2_MOUSE	O35516 mus musculus
4	169.5	2.7	2471	1 NTC2_RAT	O9QW30 rattus norv
5	166.5	2.6	1429	1 L112_CAEEL	P14585 caenorhabdi
6	166.5	2.6	2471	1 NTC2_HUMAN	Q04721 homo sapien
7	164	2.6	2531	1 NTC1_RAT	Q07008 rattus norv
8	163.5	2.6	2437	1 NTC1_BRARE	P06530 brachydanio
9	163.5	2.6	3135	1 S230_PLAFO	Q08372 plasmodium
10	162.5	2.6	1005	1 RA50_METUA	Q58718 methanococc
11	162	2.6	2531	1 NTC1_MOUSE	Q01705 mus musculus
12	158.5	2.5	1961	1 MYH9_RAT	Q62812 rattus norv
13	158	2.5	1163	1 SBCC_CLOAB	Q971K1 clostridium
14	157.5	2.5	1978	1 MYHB_CHICK	P10587 gallus gall
15	155.5	2.5	2321	1 NTC3_HUMAN	Q9UM47 homo sapien
16	155.5	2.4	1251	1 RBP2_PLAVB	Q00799 plasmodium
17	154.5	2.4	1682	1 MSP1_PLAF3	P19598 plasmodium
18	153.5	2.4	1960	1 MYH9_HUMAN	P35579 homo sapien
19	153.5	2.4	2663	1 CENE_HUMAN	Q02224 homo sapien
20	153	2.4	1169	1 SMC_METJA	Q59037 methanococc
21	153	2.4	1225	1 SMC1_YEAST	P04934 plasmodium
22	153	2.4	1726	1 MSP1_PLAFC	P04935 plasmodium
23	153	2.4	1726	1 MSP1_PLAFC	P50495 plasmodium
24	153	2.4	1964	1 NTC4_MOUSE	P31695 mus musculus
25	152.5	2.4	2003	1 NTC4_HUMAN	Q99466 homo sapien
26	152	2.4	1957	1 YD86_SCHPO	Q10411 schizosacch
27	150.5	2.4	1526	1 MY52_SCHPO	Q9US16 schizosacch
28	150.5	2.4	1630	1 MSP1_PLAFK	P04932 plasmodium
29	150.5	2.4	1639	1 MSP1_PLAFW	P04933 plasmodium
30	149	2.4	2319	1 NTC3_RAT	Q91172 rattus norv
31	148.5	2.3	2524	1 NTC3_XENLA	P21783 xenopus lae
32	148	2.3	2318	1 NTC3_MOUSE	Q61982 mus musculus
33	147	2.3	1031	1 RAD2_YEAST	P07276 saccharomyc

## ALIGNMENTS

### RESULT 1

ID	MSP1_PLAFM	STANDARD;	PRT;	1701 AA.
AC	P08569;			
DT	01-AUG-1988 (Rel. 08, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	Merozoite surface protein 1 precursor (Merozoite surface antigens)			
DE	(PMMSA) (P190).			
GN	MSP-1.			
OS	Plasmodium falciparum (isolate mad20 / Papua New Guinea).			
OC	Eukaryote; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
OX	NCBI_TaxID=70153;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=88011243; PubMed=3079521;			
RA	Tanabe K., Mackay M., Goman M., Scaife J.G.;			
RT	"Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmodium falciparum.";			
RL	J. Mol. Biol. 195:273-287(1987).			
RN	[2]			
RP	REVISIONS TO 1403; 1569 AND 1629.			
RA	Tanabe K.;			
RL	Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE OF 1-115 FROM N.A.			
RX	MEDLINE=86136024; PubMed=3004972;			
RA	Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,			
RA	Stunnenberg H., Buyard H.;			
RT	"Polymorphism of the precursor for the major surface antigens of Plasmodium falciparum merozoites: studies at the genetic level.";			
RL	EMBO J. 4:3823-3829(1985).			
CC	-1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (Potential).			
CC	-1- PFM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42 kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; X05624; CAA29112.1; -			
DR	PIR; A26868; A26868.			
DR	PIR; B25120; B25120.			
DR	InterPro; IPR000561; EGF-like.			
DR	Pfam; PF00008; EGF; 1.			
KW	Malaria; Merozoite; Polypeptide; Repeat; Signal; Glycoprotein;			
KW	Transmembrane; GPI-anchor.			
FT	SIGNAL 1 19			POTENTIAL.
FT	CHAIN 20 1701			MEROZOITE SURFACE PROTEIN 1.
FT	CARBOHYD 110 110			N-LINKED (GLCNAC. . .) (POTENTIAL).

34	146	2.3	1875	1	MLP1_YEAST	Q02455 saccharomyc
35	145.5	2.3	2095	1	RPL_MOUSE	P56716 mus musculu
36	145.5	2.3	2556	1	NTC1_HUMAN	P46531 homo sapien
37	143	2.3	1818	1	HMW2_MYCPN	P75471 mycoplasma
38	141.5	2.2	1744	1	TANA_XENLA	Q01550 xenopus lae
39	140	2.2	3685	1	DMD_HUMAN	P11532 homo sapien
40	138	2.2	1928	1	MYSL_YEAST	P08964 saccharomyc
41	137.5	2.2	908	1	ATMB_SALTY	P22036 salmonella
42	137	2.2	1328	1	YMU0_YEAST	Q04670 saccharomyc
43	137	2.2	1939	1	MYH6_MESAU	P13539 mesocricetu
44	136.5	2.2	1295	1	GLP1_CAEEL	P13508 caenorhabdi
45	136.5	2.2	1372	1	FUS1_SCHPO	Q10719 schizosacch

FT CARBOHYD 239 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 470 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 536 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 607 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 802 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 899 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 919 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 965 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 991 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1089 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1196 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1588 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 1701 AA; 193768 MW; 3FC2EC59AF6EA98 CRC64;  
Query Match 2.7%; Score 171; DB 1; Length 1701;  
Best Local Similarity 19.4%; Pred. No. 0.057;  
Matches 246; Conservative 164; Mismatches 437; Indels 420; Gaps 62;  
QY 107 EPTKSEKQELCT--HCIKVPMVLDPALPANITLKDLPSPFHSASDIFNVAKPK 164  
DB 27 ELVKKLEALEDAVLGYSLFQKEMVNLNKGSTGTAVTSTPGSSGSGVTSGGSVASVASVA 86  
QY 165 NPSTNVSVVFDSTKDVDAISGLKGNRQT-----VWRGYLTDT 205  
DB 87 SGGSGGSV-----ASGG--SGNSRRTPNSDSDSNKTYADLKHVRQNYLFTI 133  
QY 206 KEVPLGLVMDLAFSLGPPPTFKETNOLTKLPENLSKVLLQLQYSEASVALLKLNPK 265  
DB 134 KELK-----YPELFDLTHMLT--LSKNVDGFKLIDGYEINELLYKLNYY 179  
QY 266 DF--OELNK-----QTKNMTIDGKELTISPAYLLWDLISAISQSKD-----ED----- 307  
DB 180 DLLRAKLANDACANSYCQIPFNLIKIRANELDVLKIVFGYKRPDLNLDKNGVKMEDYIKKN 239  
QY 308 -----ISASRPENEE-----LRSLSRERHAPWVRNIFVTNGQ 343  
DB 240 KTTIANINELIEGSKTTIDQKNADNEEGKKLYQAQNYLFYNNKQLEFAHNLISVLEKR 299  
QY 344 IPSWLNLDNPRVTIVTHODVFNLSHLPTFSSP-----AIESHVRHREGLSKQFTY 394  
DB 300 IDTLAKNENIKLLIEDIKYTDANPTGSKPNLPENKKEVEGHEKKEIA-KTIK 358  
QY 395 LNDVFMGKDVMPDFYSHSGKQKYLT-----WVPVPCAEGCGPSWKIDG 440  
DB 359 FNIDSLF--TDPLEYLYLRKNKKVDVTPKSDPTKSVQIPKVPYPN----- 404  
QY 441 YCDKACNNSACDWDGDCSGSGSRYTAGGGTSGISGVGPQWQFGGINSVSYNCQCA 500  
DB 405 -----GIVYPLPLTDIHNLSAADND--K 425  
QY 501 NSWLADKFCDAQCNVLSGCFDAGD--CGODHFEHLYKVTLLPNQTHYITPKGCLPYFSFA 559  
DB 426 NSY-----GDLNPDTEKINEKILTDNKKKI-----FI 455  
QY 560 EVAKRGVEGAYSDNPPIIRHASTANKWTKIHLIMHSGMNATTI---HFNLTFQNTNDEE-- 614  
DB 456 NNKKOID---LEKNINHTKQNK--KLEDDYKSKDYEELEKFEYEMFNPNFDKDV 511  
QY 615 ---FKMQITVEVD--TREGPKLNSTAKGYENLVSPITLLPEAIIIFEDIPKEKRPK--F 668  
DB 512 DKIFARYTYNVEKORYNNKTFSS-----NNSVYNVQKLKALSYLEDSLRKGISERDF 566  
QY 669 KHDVNSY-----BRAQEV-----KIPLVNIISLLPKDAQSLNLTLD--LOLEHGDI 714  
DB 567 NHYITLTKGLEADIKKLEEKSSKNKLEKRNFKGLTHSANASLEVSIDVLKQVQKVLII 626  
QY 715 LKGYNLKSALLRSPLMNSHAKIKNOAITD-----ETNDSLVAPOEKQVHKSILP 766  
DB 627 KKIEDLRKIEL--FLKNAQ--LKDSIHVPNIYKPNKPEPYIYLKKEVDLKEKFP 680  
QY 767 NSLGYSER---LQRLTTPAVSVKVNHDQGNPL-----DLETTARFVETHQKTI 817  
DB 681 KYKMLKKEQAVALSSITQPLVAASETTEDGGHSTHTLSQSGTEVTEETEVEV---TEETVG 737

QY 818 --GNVTKEKPPSLIPLVLESQMTKEKKITGKKEKNSRMEENAHGIVTEVLGRKLQHY- 874  
DB 738 HTTIVTILPK-----EESAPKEVKV-----ENSIEHKSNDNSQALTKTVYLLKLDL 788  
QY 875 TDSYGLFLPWEKKYFL-----DLL-----DEESLTKQLAYFTDSKNRARYKRD 919  
DB 789 TKSYSI-----CHKYILVSNSSMDQKLELVNLTDEEEKELKS----- 825  
QY 920 TFADSLRVNKLNSKFGFTSRKYPAPHPMDRIVMQLQDMFPPEEDKTSFHKVRHSE 979  
DB 826 --CDPLDLFLNIQNN-----IPA-MYSLYDS-MNNDLQHLHFFELYOK----- 863  
QY 980 DMQFAFSYFYILMSAVQPLNISQVDE-----VDTQSGSVLSDEIR- 1021  
DB 864 -----EMYYLHLKKEENHKKLEEQKQITGTSTSPGNTVTYTAQSATHSNSQOQ 917  
QY 1022 -----TLATRIHELPLSL-----QDITGLEHMLI--NCSKM---LPAD 1054  
DB 918 SNASSTNTQNGVAVSSGPAVVEESH-PLTVLSISNDLKGIVSLLNGLNKTVPNPLTIS 976  
QY 1055 ITOL-----NNIPPTQESYDNLPPVTKSLVTKNCPVT-----DKIKKAYKDK- 1098  
DB 977 TTEMEKFEYENILKNDTYFNDDIKQFVKS---NSKVITGLTETQNALNDEI-KKLDL 1032  
QY 1099 -----NKYRFEI-----MGEETAFKMIKMTNVSHVVGOLDIRKKNPKFVCLND 1142  
DB 1033 QLSFDLYNKYKLDRLFNKKGKELGQDKMQIKKLLKKEQLESKLSL--NNPHN-VLQNF 1090  
QY 1143 NIDHNH 1149  
DB 1091 SVFENK 1097

## RESULT 2

MSPL\_PLAFF ID MSP1\_PLAFF STANDARD; PRT; 1701 AA.  
AC P13819;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)  
DE (PMWSA).  
GN MSP-1.  
OS Plasmodium falciparum (isolate FC27 / Papua New Guinea).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5837;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88142999; PubMed=2449612;  
RA Peterson M.G., Coppel R.L., McIntyre P., Langford C.J., Woodrow G.,  
Brown G.V., Anders R.F., Kemp D.J.;  
RT "Variation in the precursor to the major merozoite surface antigens  
of Plasmodium falciparum.";  
RL Mol. Biochem. Parasitol. 27:291-302(1988).  
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor  
(Potential).  
CC -!- PFM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42  
kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF  
MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DB EMBL; M19143; AAA29653.1; -.  
DB PIR; A54498; A54498.  
DB InterPro; IPR000561; EGF-like.  
DB Pfam; PF00008; EGF; 1.

KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;  
KW Transmembrane; GPI-anchor.  
FT SIGNAL 1 19 POTENTIAL.  
FT CHAIN 20 1701 MEROZOITE SURFACE PROTEIN 1.  
FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 470 470 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 536 536 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 607 607 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 802 802 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 899 899 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 919 919 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 965 965 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 991 991 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1196 1196 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1588 1588 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 1701 AA; 193719 MW; 3920875E73D38552 CRC64;  
Query Match 2.7%; Score 170; DB 1; Length 1701;  
Best Local Similarity 19.4%; Pred. No. 0.065;  
Matches 246; Conservative 164; Mismatches 437; Indels 420; Gaps 62;  
QY 107 EPTKSEKOLECLLT--HCIKVPMVLDPALPANITLKDLPSLYPSFHSASDIFENVAKPK 164  
DB 27 ELVKKLEALEDVLTGYSLSFQKEKVVLENGTSGTAVTTTPGSSGSGVTSGGSAVASVA 86  
QY 165 NPSTNVSVVDFSTKDVEDAHSGLLKSGNSROT-----ASGG--SGNSRTPNSDSSNTKYADLKHVRQNYLFTI 133  
DB 87 SGGSGGVS-----ASGG--SGNSRTPNSDSSNTKYADLKHVRQNYLFTI 133  
QY 206 KEVPLGLVQLADLAFLSGPPTKETNQLTKLPENLSSKVKLQYSEASVALLKLNPK 265  
DB 134 KELK-----YPELDTLNHMLT--LSKNVDGFKYLDIGYEEINELLYKLNFFY 179  
QY 266 DF--QELNK-----OTKKNMTIDGKELTISPAYLLWLSAISOSKOD-----ED----- 307  
DB 180 DLLRAKLNACANSYCOIFENKIRANELDVLKKVFGYRKPLDNIKNKDNVGMEDYIKKN 239  
QY 308 -----ISARFEDNEE-----LRYSLRTERRHAPWVRNFIWTNGQ 343  
DB 240 KTTIANINELIEGSKTTIDQKNADNEEGKKLYQAQYNYLQLOEHLNLSVLEKR 299  
QY 344 IPSWLNDNPRVTIVHQVFNLSHLPTFSP-----AISHVHRIBGLSOKTIY 394  
DB 300 IDTLKKNENIKKLEIDIKITDAENPTTGSKPENLPENKKKEVSGHEKIKEIA--KTIK 358  
QY 395 LNDVVMFGKVDWPDFFYSHSGKQVYLT-----WVPNCAEGCGPGSWIKDG 440  
DB 359 FNIDSLF--TDPLEYLYLEKKNKVDVTPKSDPTKSVQIPKVPYPN----- 404  
QY 441 YCDKACNNSACDWGDCSGNSGSGRYIAGGGGTSGYGVGPWFQFGGINSVSYCNQGA 500  
DB 405 -----GIVYPLPLTDIHNLSAADND--K 425  
QY 501 NSHLADKFCDAQNLSCGFDAGD--CGQDFHELYKVVILLPNQTHYIIPKGCCLPYFSA 559  
DB 426 NSY-----GDLNPDTEKINEKIITDNKERKI-----FI 455  
QY 560 EVAKRGEAGYSDNPITRIHASIANKWKTHLTHMSGMATTI---HFNLTFQNTNDEE-- 614  
DB 456 NNKKOID---LEERNINTKQNK--KLEDYEKSKDYEELEKFEKFNNNFDKDV 511  
QY 615 ----FRMOITVEVD--TREGPKLNSTAQGYENLVSPITLLPEAIFLFDIPKREPK--F 668  
DB 512 DKIFSARYTYNKEQRYNNKFESS-----NNSVYNVQKLKALSYLEDSLRKLGISEKDF 566  
QY 669 KRDVNST-----RRAQEV-----KPLVNLISLPKDAQSLNTLD---LQLEHGDIT 714  
DB 567 NHYITLKTGLEADIKKLEIKESSENKILEKFKGLTHSANASLEVSIDIVKLQKVLLI 626  
QY 715 LKGYNLKSKALLRSFLMNSQAHKIKNQAIITD-----ETNDSLVAPOEKQVHKSILP 766

DB 627 KKIEDLRKIEL---FLKNAQ---LKDSIHVPIYKPNKPEFYLIIVLKEVDKLEFIP 680  
QY 767 NSLGSYER---LORLTTPAVSVKNGHDQGNPPL-----DLETARFRVETHTQKTIG 817  
DB 681 KVKOMLKEQAVLSITOTPLVAASETTEDGGHSTHTLSQSGETEVTEVTEV---TEETVG 737  
QY 818 --GNVTKKPPSLIVPLESQMTKEKKTGKEKNSRMEENAHGIVTEVLLGRKLQHY- 874  
DB 738 HTTIVTITLPPK-----EESAPKEVKV---ENSIEHKSNDNSQALTKTVLKKLDEFL 788  
QY 875 TDSYLGFLPWKKYFL-----DLL-----DREESLKTOLAYFTDSKNRARIKRD 919  
DB 789 TKSYI-----CHKYILVSNSSMDOKLLEVYNLTPEENELKS----- 825  
QY 920 TFADSLRYVVKILNSKFGFTSRKVPAPMHPMIDRIVMQELQDMPPEFDTSPHKVRHSE 979  
DB 826 --CDPLDLLFNQNN-----IPA-MYSLYDSKNI-DLQHLFFELYQK----- 863  
QY 980 DMQFAFSYFYLYMSAVQPLNISQVDFE-----VDTQDSGLVLSDRIR- 1021  
DB 864 -----EMYYLHLKEENHIKLEEQKQITGTSTSSPGNTTNTVAQSATHSNSQNOQ 917  
QY 1022 -----TLATRIHELPLSL-----QDLTGLEHMLI--NCSKM---LPAD 1054  
DB 918 SNASSTNTQNGVAVSSGPAVVEESH-PLTVLSISNDLKGIVLSLNLGNKTKVPNPLTIS 976  
QY 1055 ITQL-----NNIPPTQESYIDPNLPVTKSLVTNCKPVT-----DKIHKAYKDK- 1098  
DB 977 TTEKEFVENLLKNDYFNDIDIKQVKS---NSKVITGLTETOKNALNDEI-KLKDITL 1032  
QY 1099 -----NKYRFEI-----MGEIEIAFKMIRTNVSHVVGQLDIRKNRKFVCLND 1142  
DB 1033 QLSFDLYNKYKLDRLFNKKELQDQMKQIKKTLLEQLESKLNSL--NNPHN-VLQNF 1090  
QY 1143 NIDHNH 1149  
DB 1091 SVFFENKK 1097  
RESULT 3  
NTC2\_MOUSE STANDARD; PRG; 2470 AA.  
ID NTC2\_MOUSE  
AC O3516; Q60941; Q60608;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Neurogenic locus notch homolog protein 2 precursor (Notch 2) (Notch B).  
GN NOTCH2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=Thymus;  
RA Hamada Y., Higuchi M., Tsujimoto Y.;  
RT "Complete amino acid sequence and multiform transcripts encoded by a single copy of mouse Notch2 gene.";  
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 316-1518 FROM N.A.  
RC STRAIN=C57BL/6 X CBA; TISSUE=Embryo;  
RX MEDLINE=93178563; PubMed=8440332;  
RA Lardelli M., Lendahl U.;  
RT "Notch A and Notch B-two mouse Notch homologues coexpressed in a wide variety of tissues.";  
RL Exp. Cell Res. 204:364-372(1993).  
RN [3]  
RP SEQUENCE OF 1765-2153 FROM N.A.  
RX MEDLINE=97075110; PubMed=8917536;  
RA Milner L.A., Bigas A., Kopan R., Brashem-Stein C., Bernstein I.D.,  
RA Martin D.I.;



FT "Inhibition of granulocytic differentiation by mNotch1.";  
RL Proc. Natl. Acad. Sci. U.S.A. 93:13014-13019(1996).  
RN [4]  
RP FUNCTION  
RX MEDLINE=99396706; PubMed=10393120;  
RA Hamada Y., Kadokawa Y., Okabe M., Ikawa M., Coleman J.R.,  
RA Tsujimoto Y.;  
RT "Mutation in ankyrin repeats of the mouse Notch2 gene induces early  
RT embryonic lethality.";  
RL Development 126:3415-3424(1999).  
RN [5]  
RP DEVELOPMENTAL STAGE, AND ALTERNATIVE SPLICING.  
RX MEDLINE=95333893; PubMed=7609614;  
RA Higuchi M., Kiyama H., Hayakawa T., Hamada Y., Tsujimoto Y.;  
RT "Differential expression of Notch1 and Notch2 in developing and adult  
RT mouse brain.";  
RL Brain Res. Mol. Brain Res. 29:263-272(1995).  
RN [6]  
RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.  
RX MEDLINE=21523956; PubMed=11518718;  
RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;  
RT "Murine notch homologs (NL-4) undergo presenilin-dependent  
RT proteolysis.";  
RL J. Biol. Chem. 276:40268-40273(2001).  
RN [7]  
RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.  
RX MEDLINE=21374376; PubMed=11459941;  
RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;  
RT "Conservation of the biochemical mechanisms of signal transduction  
RT among mammalian Notch family members.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).  
RN [8]  
RP FUNCTION: Functions as a receptor for membrane-bound ligands  
CC Jagged1, Jagged2 and Deltal to regulate cell-fate determination.  
CC Upon ligand activation through the released notch intracellular  
CC domain (NICD) it forms a transcriptional activator complex with  
CC RBP-J kappa and activates genes of the enhancer of split locus.  
CC Affects the implementation of differentiation, proliferation and  
CC apoptotic programs (By similarity). May play an essential role in  
CC postimplantation development, probably in some aspect of cell  
CC specification and/or differentiation.  
CC [9]  
CC SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-  
CC terminal fragment N(EC) which are probably linked by disulfide  
CC bonds.  
CC [10]  
CC SUBCELLULAR LOCATION: Type I membrane protein. Following  
CC proteolytical processing NICD is translocated to the nucleus.  
CC [11]  
CC ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be  
CC produced by alternative splicing.  
CC [12]  
CC TISSUE SPECIFICITY: Expressed in the brain, liver, kidney,  
CC neuropathia, somites, optic vesicles and branchial arches, but  
CC not heart.  
CC [13]  
CC DEVELOPMENTAL STAGE: Expressed in the embryonic ventricular zone,  
CC the postnatal ependymal cells, and the choroid plexus throughout  
CC embryonic and postnatal development.  
CC [14]  
CC PTM: Synthesized in the endoplasmic reticulum as an inactive form  
CC which is proteolytically cleaved by a furin-like convertase in the  
CC trans-Golgi network before it reaches the plasma membrane to yield  
CC an active, ligand-accessible form. Cleavage results in a C-  
CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following  
CC ligand binding, it is cleaved by TNF-alpha converting enzyme  
CC (TACE) to yield a membrane-associated intermediate fragment called  
CC notch extracellular truncation (NEXT). This fragment is then  
CC cleaved by presenilin dependent gamma-secretase to release a  
CC notch-derived peptide containing the intracellular domain (NICD)  
CC from the membrane.  
CC [15]  
CC PTM: Phosphorylated.  
CC [16]  
CC SIMILARITY: BELONGS TO THE NOTCH FAMILY.  
CC [17]  
CC SIMILARITY: CONTAINS 34.5 EGF-LIKE DOMAINS.  
CC [18]  
CC SIMILARITY: CONTAINS 2 LIN/NOTCH REPEATS.  
CC [19]  
CC SIMILARITY: CONTAINS 6 ANK REPEATS.  
CC [20]  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration -  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; D32210; BAA22094.1; -;  
CC EMBL; X68279; CAA48340.1; -;  
CC EMBL; U31881; AAC52924.1; -;  
CC HSSP; P16109; 1FSB.  
CC MGD; MGI:97364; Notch2.  
CC InterPro; IPR002110; ANK.  
CC InterPro; IPR000152; Asx\_hydroxyl.  
CC InterPro; IPR000561; EGF-like.  
CC InterPro; IPR000742; EGF-2.  
CC InterPro; IPR001881; EGF-Ca.  
CC InterPro; IPR001438; EGF-II.  
CC InterPro; IPR000800; Notch.  
CC Pfam; PF00008; EGF; 35.  
CC Pfam; PF00023; ank; 6.  
CC Pfam; PF00066; notch; 2.  
CC PRINTS; PR00010; EGFBLD.  
CC PRINTS; PR01452; NOTCH.  
CC SMART; SM00248; ANK; 4.  
CC SMART; SM00179; EGF-Ca; 22.  
CC SMART; SM00001; EGF-like; 12.  
CC SMART; SM00004; NL; 3.  
CC PROSITE; PS00088; ANK\_REPEAT; 4.  
CC PROSITE; PS0297; ANK\_REPEAT\_REGION; 1.  
CC PROSITE; PS00010; ASX\_HYDROXYL; 22.  
CC PROSITE; PS00022; EGF-1; 33.  
CC PROSITE; PS01186; EGF-2; 27.  
CC PROSITE; PS01187; EGF-Ca; 22.  
CC Receptor; Transcription regulation; Activator; Differentiation;  
CC Developmental protein; Repeat; ANK repeat; EGF-like domain;  
CC Transmembrane; Glycoprotein; Signal; Phosphorylation;  
CC Alternative splicing.  
CC SIGNAL 1 25  
CC CHAIN 26 2470 POTENTIAL.  
CC CHAIN 1666 2470 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 2.  
CC CHAIN 1697 2470 NOTCH EXTRACELLULAR TRUNCATION.  
CC DOMAIN 26 1677 NOTCH INTRACELLULAR DOMAIN.  
CC TRANSMEM 1678 1698 EXTRACELLULAR (POTENTIAL).  
CC DOMAIN 1699 2470 POTENTIAL.  
CC DOMAIN 26 63 CYTOPLASMIC (POTENTIAL).  
CC DOMAIN 105 143 EGF-LIKE 1.  
CC DOMAIN 144 180 EGF-LIKE 2.  
CC DOMAIN 182 219 EGF-LIKE 3.  
CC DOMAIN 221 256 EGF-LIKE 4.  
CC DOMAIN 236 334 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).  
CC DOMAIN 336 372 EGF-LIKE 6 (INCOMPLETE).  
CC DOMAIN 373 411 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).  
CC DOMAIN 413 452 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).  
CC DOMAIN 454 490 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).  
CC DOMAIN 492 528 EGF-LIKE 10.  
CC DOMAIN 530 566 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).  
CC DOMAIN 568 603 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).  
CC DOMAIN 605 641 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).  
CC DOMAIN 643 678 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).  
CC DOMAIN 680 716 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).  
CC DOMAIN 718 753 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).  
CC DOMAIN 755 791 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).  
CC DOMAIN 793 829 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).  
CC DOMAIN 831 869 EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).  
CC DOMAIN 871 907 EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).  
CC DOMAIN 909 945 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).  
CC DOMAIN 947 983 EGF-LIKE 22, CALCIUM-BINDING (POTENTIAL).  
CC DOMAIN 985 1021 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).  
CC DOMAIN 1023 1059 EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).  
CC DOMAIN 1061 1097 EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).  
CC DOMAIN 1099 1145 EGF-LIKE 26, CALCIUM-BINDING (POTENTIAL).  
CC DOMAIN 1147 1183 EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).  
CC EGF-LIKE 28, CALCIUM-BINDING (POTENTIAL).  
CC EGF-LIKE 29, CALCIUM-BINDING (POTENTIAL).  
CC EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).



FT	CHAIN	1697	2471	SIMILARITY).	FT	DISULFID	286	295	BY SIMILARITY.
FT	DOMAIN	26	1677	NOTCH INTRACELLULAR DOMAIN (BY	FT	DISULFID	302	315	BY SIMILARITY.
FT	DOMAIN	1678	1698	SIMILARITY).	FT	DISULFID	309	324	BY SIMILARITY.
FT	DOMAIN	1699	2471	EXTRACELLULAR (POTENTIAL).	FT	DISULFID	326	335	BY SIMILARITY.
FT	DOMAIN	26	63	POTENTIAL.	FT	DISULFID	342	353	BY SIMILARITY.
FT	DOMAIN	64	102	CYTOPLASMIC (POTENTIAL).	FT	DISULFID	347	362	BY SIMILARITY.
FT	DOMAIN	64	102	EGF-LIKE 1.	FT	DISULFID	364	373	BY SIMILARITY.
FT	DOMAIN	105	143	EGF-LIKE 2.	FT	DISULFID	379	390	BY SIMILARITY.
FT	DOMAIN	144	180	EGF-LIKE 3.	FT	DISULFID	384	401	BY SIMILARITY.
FT	DOMAIN	182	219	EGF-LIKE 4.	FT	DISULFID	403	412	BY SIMILARITY.
FT	DOMAIN	221	258	EGF-LIKE 5.	FT	DISULFID	419	433	BY SIMILARITY.
FT	DOMAIN	260	296	EGF-LIKE 6.	FT	DISULFID	427	442	BY SIMILARITY.
FT	DOMAIN	298	336	EGF-LIKE 7.	FT	DISULFID	444	453	BY SIMILARITY.
FT	DOMAIN	338	374	EGF-LIKE 8.	FT	DISULFID	460	471	BY SIMILARITY.
FT	DOMAIN	375	413	EGF-LIKE 9.	FT	DISULFID	465	480	BY SIMILARITY.
FT	DOMAIN	415	454	EGF-LIKE 10.	FT	DISULFID	482	491	BY SIMILARITY.
FT	DOMAIN	456	492	EGF-LIKE 11.	FT	DISULFID	498	509	BY SIMILARITY.
FT	DOMAIN	494	530	EGF-LIKE 12.	FT	DISULFID	503	518	BY SIMILARITY.
FT	DOMAIN	532	568	EGF-LIKE 13.	FT	DISULFID	520	529	BY SIMILARITY.
FT	DOMAIN	570	605	EGF-LIKE 14.	FT	DISULFID	536	547	BY SIMILARITY.
FT	DOMAIN	607	643	EGF-LIKE 15.	FT	DISULFID	541	556	BY SIMILARITY.
FT	DOMAIN	645	680	EGF-LIKE 16.	FT	DISULFID	558	567	BY SIMILARITY.
FT	DOMAIN	682	718	EGF-LIKE 17.	FT	DISULFID	574	584	BY SIMILARITY.
FT	DOMAIN	720	755	EGF-LIKE 18.					
FT	DOMAIN	757	793	EGF-LIKE 19.					
FT	DOMAIN	795	831	EGF-LIKE 20.					
FT	DOMAIN	833	871	EGF-LIKE 21.					
FT	DOMAIN	873	909	EGF-LIKE 22.					
FT	DOMAIN	911	947	EGF-LIKE 23.					
FT	DOMAIN	949	985	EGF-LIKE 24.					
FT	DOMAIN	987	1023	EGF-LIKE 25.					
FT	DOMAIN	1025	1061	EGF-LIKE 26.					
FT	DOMAIN	1063	1099	EGF-LIKE 27.					
FT	DOMAIN	1101	1147	EGF-LIKE 28.					
FT	DOMAIN	1149	1185	EGF-LIKE 29.					
FT	DOMAIN	1187	1223	EGF-LIKE 30.					
FT	DOMAIN	1225	1262	EGF-LIKE 31.					
FT	DOMAIN	1264	1302	EGF-LIKE 32.					
FT	DOMAIN	1304	1343	EGF-LIKE 33.					
FT	DOMAIN	1374	1412	EGF-LIKE 34.					
FT	DOMAIN	1645	1648	EGF-LIKE 35.					
FT	DOMAIN	1994	1997	POLY-ALA.					
FT	DOMAIN	2426	2429	POLY-LEU.					
FT	DOMAIN	2446	2451	POLY-SER.					
FT	DOMAIN	1420	1456	POLY-GLY.					
FT	REPEAT	1503	1535	LIN/NOTCH 1.					
FT	REPEAT	1827	1871	LIN/NOTCH 2.					
FT	REPEAT	1876	1905	ANK 1.					
FT	REPEAT	1909	1939	ANK 2.					
FT	REPEAT	1943	1972	ANK 3.					
FT	REPEAT	1976	2005	ANK 4.					
FT	REPEAT	2009	2038	ANK 5.					
FT	REPEAT	28	41	ANK 6.					
FT	DISULFID	28	41	BY SIMILARITY.					
FT	DISULFID	35	51	BY SIMILARITY.					
FT	DISULFID	53	62	BY SIMILARITY.					
FT	DISULFID	68	79	BY SIMILARITY.					
FT	DISULFID	73	90	BY SIMILARITY.					
FT	DISULFID	92	101	BY SIMILARITY.					
FT	DISULFID	109	121	BY SIMILARITY.					
FT	DISULFID	115	131	BY SIMILARITY.					
FT	DISULFID	133	142	BY SIMILARITY.					
FT	DISULFID	148	159	BY SIMILARITY.					
FT	DISULFID	153	168	BY SIMILARITY.					
FT	DISULFID	170	179	BY SIMILARITY.					
FT	DISULFID	186	198	BY SIMILARITY.					
FT	DISULFID	192	207	BY SIMILARITY.					
FT	DISULFID	209	218	BY SIMILARITY.					
FT	DISULFID	225	236	BY SIMILARITY.					
FT	DISULFID	230	246	BY SIMILARITY.					
FT	DISULFID	248	257	BY SIMILARITY.					
FT	DISULFID	264	275	BY SIMILARITY.					
FT	DISULFID	269	284	BY SIMILARITY.					

Query Match 2.7%; Score 169.5; DB 1; Length 2471;

Best Local Similarity 26.2%; Pred. No. 0.12;

Matches 44; Conservative 15; Mismatches 58; Indels 51; Gaps 5;

QY 407 PDFYSHKQKQVYTWVPNCAEGCPGSKWTKDGYCDKACNNSACDWDGDCSGSGSR 466

DB 1404 PPFWSHCSYTPATSTPTATCLSQYCADKARDGICDEACNSHACQWDGDC----- 1455

QY 467 YIAGGGTSGIGGQPW-QFGGGINSVYCGQG-----AN 501

DB 1456 -----SLTWEDPWANCTSSLRWEYINNQDELONTAECLFDFECORNSKTKYD 1506

QY 502 SWLADK-----CDQACNVLSGFDAGDCGDHFLHYK-----VILLP 540

DB 1507 KYCADHFNDHCKGNCNNECGWGLDCAADQENLAEGILVIVVLLP 1554

## RESULT 5

L112-CAEEL

ID L112-CAEEL STANDARD; PRT; 1429 AA.

AC P14585;

DT 01-JAN-1990 (Rel. 13, Created)

DT 01-JAN-1990 (Rel. 13, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Lin-12 protein precursor.

GN LIN-12 OR R107.8.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI\_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-Bristol N2;

RX MEDLINE-88334747; PubMed=3419531;

RA Yochem J., Weston K., Greenwald I.;

RT "The Caenorhabditis elegans lin-12 gene encodes a transmembrane

RL Nature 335:547-550(1988).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-Bristol N2;

RX MEDLINE-94150718; PubMed=7906398;

RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,

RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,

RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,

RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,

RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,

RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M., R.,

RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen M., R.,

RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,  
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,  
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,  
RA Wohldman P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
elegans.";  
RL Nature 368:32-38(1994).  
CC -!- FUNCTION: LIN-12 IS INVOLVED IN SEVERAL CELL FATES DECISIONS THAT  
CC REQUIRES CELL-CELL INTERACTIONS. IT IS POSSIBLE THAT LIN-12  
CC ENCODES A MEMBRANE-BOUND RECEPTOR FOR A SIGNAL THAT ENABLES  
CC EXPRESSION OF THE VENTRAL UTERINE PRECURSOR CELL FATE.  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -!- SIMILARITY: HIGH, TO C.ELEGANS GLP-1.  
CC -!- SIMILARITY: CONTAINS 13 EGF-LIKE DOMAINS.  
CC -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.  
CC -!- SIMILARITY: CONTAINS 5 ANK REPEATS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M12069; AAA70191.1; -;  
DR EMBL; 214092; CAA78474.1; -;  
DR PIR; S06434; S06434..  
DR HSP; P00740; IEDM.  
DR WormPep; R107.8; CE00274.  
DR InterPro; IPR002110; ANK.  
DR InterPro; IPR000152; Asx\_hydroxyl.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR000742; EGF\_2.  
DR InterPro; IPR001881; EGF\_Ca.  
DR InterPro; IPR000800; Notch.  
DR Pfam; PF00008; EGF; 13.  
DR Pfam; PF00023; ank; 6.  
DR Pfam; PF00066; notch; 3.  
DR PRINTS; PR01452; NOTCH.  
DR SMART; SM00248; ANK; 4.  
DR SMART; SM00179; EGF\_CA; 2.  
DR SMART; SM00001; EGF\_like; 10.  
DR SMART; SM00004; NL; 2.  
DR PROSITE; PS50088; ANK\_REPEAT; 3.  
DR PROSITE; PS50297; ANK\_REPEAT\_REGION; 1.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 3.  
DR PROSITE; PS00022; EGF\_1; 12.  
DR PROSITE; PS01186; EGF\_2; 11.  
DR PROSITE; PS01187; EGF\_CA; 2.  
KW Differentiation; Repeat; ANK repeat; EGF-like domain; Transmembrane;  
KW Glycoprotein; Signal.  
FT SIGNAL 1 15  
FT CHAIN 16 1429  
FT DOMAIN 16 908  
FT TRANSMEM 909 931  
FT DOMAIN 932 1429  
FT DOMAIN 20 61  
FT DOMAIN 114 150  
FT DOMAIN 152 190  
FT DOMAIN 201 246  
FT DOMAIN 250 285  
FT DOMAIN 287 323  
FT DOMAIN 323 363  
FT DOMAIN 365 402  
FT DOMAIN 404 441  
FT DOMAIN 449 492  
FT DOMAIN 503 541  
FT DOMAIN 543 579  
FT DOMAIN 582 619  
FT REPEAT 635 669  
FT REPEAT 670 710  
FT REPEAT 711 750

FT REPEAT 1093 1122 ANK 1.  
FT REPEAT 1126 1158 ANK 2.  
FT REPEAT 1162 1194 ANK 3.  
FT REPEAT 1206 1236 ANK 4.  
FT REPEAT 1240 1269 ANK 5.  
FT DISULFID 24 35  
FT DISULFID 29 49  
FT DISULFID 51 60  
FT DISULFID 118 129  
FT DISULFID 123 138  
FT DISULFID 140 149  
FT DISULFID 156 169  
FT DISULFID 163 178  
FT DISULFID 180 189  
FT DISULFID 205 227  
FT DISULFID 221 234  
FT DISULFID 236 245  
FT DISULFID 254 264  
FT DISULFID 259 273  
FT DISULFID 275 284  
FT DISULFID 291 302  
FT DISULFID 296 311  
FT DISULFID 313 322  
FT DISULFID 327 339  
FT DISULFID 334 351  
FT DISULFID 353 362  
FT DISULFID 369 381  
FT DISULFID 375 390  
FT DISULFID 392 401  
FT DISULFID 408 419  
FT DISULFID 413 429  
FT DISULFID 431 440  
FT DISULFID 507 518  
FT DISULFID 512 529  
FT DISULFID 531 540  
FT DISULFID 547 558  
FT DISULFID 552 567  
FT DISULFID 569 578  
FT DISULFID 586 597  
FT DISULFID 591 607  
FT DISULFID 609 618  
FT CARBOHYD 41 41  
FT CARBOHYD 165 165  
FT CARBOHYD 194 194  
FT CARBOHYD 378 378  
FT CARBOHYD 515 515  
FT CARBOHYD 623 623  
FT CARBOHYD 751 751  
FT CARBOHYD 754 754  
FT CARBOHYD 900 900  
SQ SEQUENCE 1429 AA; 157115 MW; 2355EDD7A62C025DB CRC64;  
Query Match 2.6%; Score 166.5; DB 1; Length 1429;  
Best Local Similarity 19.8%; Pred. No. 0.08;  
Matches 111; Conservative 43; Mismatches 166; Indels 241; Gaps 21;  
Qy 428 CAEGCPGSGWTKDGYCDKACNNACDWGDCSGNSGSRVYAGGGTSGVGPWQ--- 484  
Db 643 CSERA-----NDGNCADACNYACKFDGDCSGKR-----EPFSKCR 679  
Qy 485 -----FGGINSVSYCNQGCAN-----SWLADKFC 510  
Db 680 YGNMCADFFANGV-----CNOACNNECLYDGMDCLPVVRCPVKTRHCASFANGIC 734  
Qy 511 QACNVLCGFDAGDCGDHFEHLYKVLLPNQTHYIIPKGECLPYFSFAEVAKRGEVAY 570  
Db 735 PECNTNCGFDGDC-----DNET----- 753  
Qy 571 SDNPPIIRHASTANKWKTHILHIMSGMNATTI-HFNLTFTONTDEEFK----- 616  
Db 754 -----NATIITNIRITVQ-MDPKEFQVGTGQSILMEISSA 786  
Qy 617 MQITVEVD-TREGPKLNSTAQKGVENLVSPITL-----LPEAEILFEDIPEKRPFKRH 671

```

Db 787 LRVTVRIORDEGLPV---FQWNGESEMVRKMKNERQLTEQHVLTSTISR-----KIKRS 838
Qy 672 DVNSTRQAQEEKIPLVNISLLPKDAQLSLNTLDLQLEHGDIITLKGYNLS-----KS 723
Db 839 ATNIGVVVILEVQENCDTGKCLYKDAQSVVDSISARLAKKGDIDFGIPISALVAEPRKS 898
Qy 724 ALLRSFL-----MNSQHAQIKRN 740
Db 899 GNNTGFLSNALLIGAGCLIVVMVLGALPGNRTKRKRMINASVYMPMEENEKRNKN 958
Qy 741 QAITDETNDLSVAPQEKV-----HKSILPNSLGHSE-----RLQRLTFP 781
Db 959 HQSITSSQHSLLSEASYDGYIKRQNELQHYSLYPNPQGYGNGNDFLGDFNHTNLQIPTPEP 1018
Qy 782 AVSVKVNHGDPQGNPLDLETTARFVE-----THTQKT---TGGNVTKEKPSLSLIVPLES 834
Db 1019 EPESPILKHTAAGSVAITEPTITRESVNIIDPRHRTVLTWLTASNSAEKSEDLIV---- 1074
Qy 835 QMTKEKKITGKEKENSMEEN 855
Db 1075 HEAKECIAAGADVNMDCDEN 1095

RESULT 6
NTC2_HUMAN STANDARD; PRT; 2471 AA.
AC Q04721; Q9H240; Q99734;
DT 15-JUN-2002 (Rel. 41, Created)
DI 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Neurogenic locus notch homolog protein 2 precursor (Notch 2) (hn2).
GN NOTCH2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

[1]
SEQUENCE FROM N.A.
TISSUE=Brain;
RA Blaumüller C.M., Mann R.S.;
RT "Complete human notch 2 (hn2) cDNA sequence";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE FROM N.A.
TISSUE=Breast tumor;
RA Correa R.G., Camargo A.A., Moreira E.S., Simpson A.J.;
RT "Human Notch2, a novel member of cell-fate determining NOTCH family.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
[3]
SEQUENCE OF 967-1229 FROM N.A.
TISSUE=T-cell;
RA Lemasson I., Devaux C., Mesnard J.M.;
RT "Partial sequence of EGF-like repeat domain of human Notch2 mRNA.";
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
[4]
SEQUENCE OF 1810-2447 FROM N.A.
TISSUE=Brain;
RX MEDLINE=93265135; PubMed=1303260;
RA Stifani S., Blaumüller C.M., Redhead N.J., Hill R.E.,
RA Artavanis-Tsakonas S.;
RT "Human homologs of a Drosophila enhancer of split gene product define a novel family of nuclear proteins.";
RL Nat. Genet. 2:119-127(1992).
[5]
POST-TRANSLATIONAL PROCESSING.
RX MEDLINE=97386453; PubMed=9244302;
RA Blaumüller C.M., Qi H., Zagouras P., Artavanis-Tsakonas S.;
RT "Intracellular cleavage of Notch leads to a heterodimeric receptor on the plasma membrane.";
RL Cell 90:281-291(1997).
[6]

```

```

RP IDENTIFICATION OF LIGANDS.
RX MEDLINE=99180765; PubMed=10079256;
RA Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangiu M.-L.,
RA Banks A., Lelman J., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;
RT "Human ligands of the Notch receptor.";
RL Am. J. Pathol. 154:785-794(1999).
CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
CC Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
CC Upon ligand activation through the released notch intracellular
CC domain (NICD) it forms a transcriptional activator complex with
CC RBP-J kappa and activates genes of the enhancer of split locus.
CC Affects the implementation of differentiation, proliferation and
CC apoptotic programs (By similarity).
CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
CC terminal fragment N(EC) which are probably linked by disulfide
CC bonds (By similarity).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC proteolytical processing NICD is translocated to the nucleus.
CC -!- TISSUE SPECIFICITY: Expressed in the brain, heart, kidney, lung,
CC skeletal muscle and liver.
CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC which is proteolytically cleaved by a furin-like convertase in the
CC trans-Golgi network before it reaches the plasma membrane to yield
CC an active, ligand-accessible form. Cleavage results in a C-
CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
CC ligand binding, it is cleaved by TMP-alpha converting enzyme
CC (TACE) to yield a membrane-associated intermediate fragment called
CC notch extracellular truncation (NEXT). This fragment is then
CC cleaved by presenilin dependent gamma-secretase to release a
CC notch-derived peptide containing the intracellular domain (NICD)
CC from the membrane (By similarity).
CC -!- PTM: Phosphorylated (By similarity).
CC -!- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
CC -!- SIMILARITY: CONTAINS 35 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 LIN/NOTCH REPEATS.
CC -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF308601; AAA36377.2; -
DR EMBL; AF315356; AAG37073.1; -
DR EMBL; U77493; AAB19224.1; -
DR HSSP; P00740; IEDM.
DR Genew; HGNC:7882; NOTCH2.
DR MIN; 600275; -
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF-II.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR000800; Notch.
DR Pfam; PF00008; EGF; 35.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00066; notch; 2.
DR PRINTS; PR00010; EGFBL00D.
DR PRINTS; PR00011; EGF_LAMININ.
DR SMART; SM00248; ANK; 4.
DR SMART; SM00179; EGF_CA; 22.
DR SMART; SM00001; EGF_Like; 12.
DR SMART; SM00004; NL; 2.
DR PROSITE; PS50088; ANK_REPEAT; 4.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 22.
DR PROSITE; PS00022; EGF_1; 34.
DR PROSITE; PS01186; EGF_2; 29.

```

DR PROSITE; PS01187; EGF\_CA; 22.  
 KW Receptor; Transcription regulation; Activator; Differentiation;  
 KW Developmental protein; Repeat; ANK repeat; EGF-like domain;  
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation.  
 FT SIGNAL 1 25  
 FT CHAIN 26 2471  
 FT CHAIN 1666 2471  
 FT CHAIN 1697 2471  
 FT CHAIN 1677 2471  
 FT DOMAIN 26 1677  
 FT TRANSMEM 1678 1698  
 FT DOMAIN 1699 2471  
 FT DOMAIN 26 63  
 FT DOMAIN 64 102  
 FT DOMAIN 105 143  
 FT DOMAIN 144 180  
 FT DOMAIN 182 219  
 FT DOMAIN 221 258  
 FT DOMAIN 260 296  
 FT DOMAIN 298 336  
 FT DOMAIN 338 374  
 FT DOMAIN 375 413  
 FT DOMAIN 415 454  
 FT DOMAIN 456 492  
 FT DOMAIN 494 530  
 FT DOMAIN 532 568  
 FT DOMAIN 570 605  
 FT DOMAIN 607 643  
 FT DOMAIN 645 680  
 FT DOMAIN 682 718  
 FT DOMAIN 720 755  
 FT DOMAIN 757 793  
 FT DOMAIN 795 831  
 FT DOMAIN 833 871  
 FT DOMAIN 873 909  
 FT DOMAIN 911 947  
 FT DOMAIN 949 985  
 FT DOMAIN 987 1023  
 FT DOMAIN 1025 1061  
 FT DOMAIN 1063 1099  
 FT DOMAIN 1101 1147  
 FT DOMAIN 1149 1185  
 FT DOMAIN 1187 1223  
 FT DOMAIN 1225 1262  
 FT DOMAIN 1264 1302  
 FT DOMAIN 1304 1343  
 FT DOMAIN 1374 1412  
 FT REPEAT 1420 1456  
 FT REPEAT 1503 1535  
 FT REPEAT 1827 1871  
 FT REPEAT 1876 1905  
 FT REPEAT 1909 1939  
 FT REPEAT 1943 1972  
 FT REPEAT 1976 2005  
 FT REPEAT 2009 2038  
 FT DOMAIN 1645 1648  
 FT DOMAIN 1994 1997  
 FT DOMAIN 2426 2429  
 FT DISULFID 28 41  
 FT DISULFID 35 51  
 FT DISULFID 53 62  
 FT DISULFID 68 79  
 FT DISULFID 73 90  
 FT DISULFID 92 101  
 FT DISULFID 109 121  
 FT DISULFID 115 131  
 FT DISULFID 133 142  
 FT DISULFID 148 159  
 FT DISULFID 153 168  
 FT DISULFID 170 179  
 FT DISULFID 186 198  
 FT DISULFID 192 207

FT DISULFID 209 218 BY SIMILARITY.  
 FT DISULFID 225 236 BY SIMILARITY.  
 FT DISULFID 230 246 BY SIMILARITY.  
 FT DISULFID 248 257 BY SIMILARITY.  
 FT DISULFID 264 275 BY SIMILARITY.  
 FT DISULFID 269 284 BY SIMILARITY.  
 FT DISULFID 286 295 BY SIMILARITY.  
 Query Match 2.68; Score 166.5; DB 1; Length 2471;  
 Best Local Similarity 19.7%; Pred. No. 0.18;  
 Matches 110; Conservative 64; Mismatches 176; Indels 207; Gaps 26;  
 QY 424 PVPNCAECPCGSKWIKDGYCDKACNNSACDWDGDCSGNSGSRVAGGGTSGVGPWP 483  
 DB 1421 PPATCLSQYCADKARDGVCDACNSHACQWDGDC-----SUTMENPW 1463  
 QY 484 -QFGGINSVSYCQGC-----ANS-----WLADKF-----CDQACN 514  
 DB 1464 ANCSPLPCWDYINNQCDELNTVECLPDNFECOGNSKTKYKCYADHFKDNHCNOGN 1523  
 QY 515 VLSGCFDAGCGQDHFEHLYK-----VILLPNOTHYIIPKGECLPYFSFAEV----- 561  
 DB 1524 SEECGWDGLDCAADQCPENLAEGTLVIVVLP-----PEQLQDARSFLRALTLLHTN 1576  
 QY 562 -AKRGVGEAYSDNPIIRHASIANKWKTIHLIMHSGMNATIHFNLTFTONTDEEKF--- 616  
 DB 1577 LRIRDSOGELMVYIYGEKSAAMKQ-----RMTRRSLPGQEQEVA 1619  
 QY 617 -MQITVEVDTRGPKLNAQYENLVSPITLLPEAILFEDIPKEKRPFKRRHDVNS 675  
 DB 1620 GSKVLEIDNRQ---CQVDSDFCFKNTDAAALLAS-----HAIOG 1657  
 QY 676 TRRAQEEVKIPVNI---SLLPKDAQSLNTLDLQ-----EHGDI 714  
 DB 1658 T-----LSYPLSVVSESLTPERTQL-LYLLAVAVIILFILLGVIMAKRKHGSLW 1710  
 QY 715 L-KGVNLSKALLRSFLMNSOHAK-----IKNOAITDETNDLSLVAPOEKQVHS 763  
 DB 1711 LPEGFTLRDA-----SNHKRRPVGQAVGLKNLSVQVSEAN--LI----- 1750  
 QY 764 ILPNSLGSYERLQRTFPFPAVSVKVNHGDDG---QNPPLDLTETARFRVETHTKTIGGN 819  
 DB 1751 ---GTGTFSEHWVDEGQPK-KVKADEALLSEEDPIDR-----RPTQOHL 1757  
 QY 820 VTKEKPPSLVPLESOMTKE-----KKTGKEKENSMEENAEH---H 859  
 DB 1798 DIRRTPSLALTPQAEQEVLDVNVVRGPDGCTPLMLASLRGSSDLSEDEDAEDSSAN 1857  
 QY 860 IGVTEVLLGRKLOHYTD 876  
 DB 1858 IITDLIVQGASLQAQTD 1874  
 RESULT 7  
 NTCL\_RAT  
 ID NTCL\_RAT STANDARD; PRT; 2531 AA.  
 AC Q07008;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Neurogenic locus notch homolog protein 1 precursor (Notch 1).  
 GN NOTCH1  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Schwann cell;  
 RX MEDLINE=92111383; PubMed=1764995;  
 RA Weinmaster G., Roberts V.J., Lemke G.;  
 RT "A homolog of Drosophila Notch expressed during mammalian  
 development.";



Development 113:199-205(1991).  
[2] CC  
RP REVISIONS TO 1652-1653.  
RA Weinmaster G.;  
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
[3] CC  
RN FUNCTION.  
RP MEDLINE=21094508; PubMed=11182080;  
RA Tanigaki K., Nogaki F., Takahashi J., Tashiro K., Kurooka H.,  
RA Honjo T.;  
RT "Notch1 and Notch3 instructively restrict bFGF-responsive multipotent  
RI neural progenitor cells to an astroglial fate.";  
RL Neuron 29:45-55(2001).  
[4] CC  
RN TISSUE SPECIFICITY.  
RP MEDLINE=93202015; PubMed=1295745;  
RA Weinmaster G., Roberts V.J., Lenke G.;  
RT "Notch2: a second mammalian Notch gene.";  
RL Development 116:931-941(1992).  
[5] CC  
RN TISSUE SPECIFICITY.  
RP MEDLINE=21331789; PubMed=11438922;  
RA Irvin D.K., Zurcher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;  
RT "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple  
RI functional roles for the Notch-DSL signaling system during brain  
development.";  
RL J. Comp. Neurol. 436:167-181(2001).  
CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands  
CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.  
CC Upon ligand activation through the released notch intracellular  
CC domain (NICD) it forms a transcriptional activator complex with  
CC RBP-J kappa and activates genes of the enhancer of split locus.  
CC Affects the implementation of differentiation, proliferation and  
CC apoptotic programs (By similarity). Acts instructively to control  
CC the cell fate determination of CNS multipotent progenitor cells,  
CC resulting in astroglial induction and neuron/oligodendrocyte  
CC suppression.  
CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-  
CC terminal fragment N(EC) which are probably linked by disulfide  
CC bonds (By similarity).  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following  
CC proteolytical processing NICD is translocated to the nucleus (By  
CC similarity).  
CC -!- TISSUE SPECIFICITY: Expressed in the brain, kidney and spleen.  
CC Expressed in postnatal central nervous system (CNS) germinal zones  
CC and, in early postnatal life, within numerous cells throughout the  
CC CNS. Found in both subventricular and ventricular germinal zones.  
CC -!- DEVELOPMENTAL STAGE: In the embryo, highest levels occur between  
CC days 12 and 14 and decrease rapidly to much lower levels in the  
CC adult.  
CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form  
CC which is proteolytically cleaved by a furin-like convertase in the  
CC trans-Golgi network before it reaches the plasma membrane to yield  
CC an active, ligand-accessible form. Cleavage results in a C-  
CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following  
CC ligand binding, it is cleaved by TNF-alpha converting enzyme  
CC (FACE) to yield a membrane-associated intermediate fragment called  
CC notch extracellular truncation (NEXT). This fragment is then  
CC cleaved by presenilin dependent gamma-secretase to release a  
CC notch-derived peptide containing the intracellular domain (NICD)  
CC from the membrane (By similarity).  
CC -!- PTM: Phosphorylated (By similarity).  
CC -!- SIMILARITY: BELONGS TO THE NOTCH FAMILY.  
CC -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.  
CC -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.  
CC -!- SIMILARITY: CONTAINS 5 ANK REPEATS.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL: X57405; CAA040667.1; -.  
CC HSSP: P00740; 1EDM.  
DR InterPro: IPR002110; ANK.  
DR InterPro: IPR000152; Asx\_hydroxyl.  
DR InterPro: IPR000561; EGF-like.  
DR InterPro: IPR000742; EGF 2.  
DR InterPro: IPR001881; EGF Ca.  
DR InterPro: IPR001438; EGF II.  
DR InterPro: IPR002049; Laminin\_EGF.  
DR InterPro: IPR000800; Notch.  
DR Pfam: PF00008; EGF; 36.  
DR Pfam: PF00023; ank; 6.  
DR Pfam: PF00066; notch; 3.  
DR PRINTS: PR00010; EGFBL00D.  
DR PRINTS: PR00011; EGFAMININ.  
DR PRINTS: PR01452; NOTCH.  
DR SMART: SM00248; ANK; 5.  
DR SMART: SM00179; EGF\_CA; 25.  
DR SMART: SM00001; EGF\_Like; 10.  
DR SMART: SM00004; NL; 2.  
DR PROSITE: PS50088; ANK\_REPEAT; 4.  
DR PROSITE: PS50297; ANK\_REPEAT\_REGION; 1.  
DR PROSITE: PS00010; ASX\_HYDROXYL; 22.  
DR PROSITE: PS00022; EGF\_1; 35.  
DR PROSITE: PS01186; EGF\_2; 26.  
DR PROSITE: PS01187; EGF\_CA; 21.  
KW Receptor; Transcription regulation; Activator; Differentiation;  
KW Developmental protein; Repeat; ANK repeat; EGF-like domain;  
KW Transmembrane; Glycoprotein; Signal; Phosphorylation.  
FT SIGNAL 1 18  
FT CHAIN 19 2531 NOTCH EXTRACELLULAR TRUNCATION (BY  
FT CHAIN 1711 2531 SIMILARITY).  
FT CHAIN 1744 2531 NOTCH INTRACELLULAR DOMAIN (BY  
FT SIMILARITY).  
FT DOMAIN 19 1723 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 1724 1746 POTENTIAL.  
FT DOMAIN 1747 2531 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 20 58 EGF-LIKE 1.  
FT DOMAIN 59 99 EGF-LIKE 2.  
FT DOMAIN 102 139 EGF-LIKE 3.  
FT DOMAIN 140 176 EGF-LIKE 4.  
FT DOMAIN 178 216 EGF-LIKE 5.  
FT DOMAIN 218 255 EGF-LIKE 6.  
FT DOMAIN 257 293 EGF-LIKE 7.  
FT DOMAIN 295 333 EGF-LIKE 8.  
FT DOMAIN 335 371 EGF-LIKE 9.  
FT DOMAIN 372 410 EGF-LIKE 10.  
FT DOMAIN 412 450 EGF-LIKE 11.  
FT DOMAIN 452 488 EGF-LIKE 12.  
FT DOMAIN 490 526 EGF-LIKE 13.  
FT DOMAIN 528 564 EGF-LIKE 14.  
FT DOMAIN 566 601 EGF-LIKE 15.  
FT DOMAIN 603 639 EGF-LIKE 16.  
FT DOMAIN 641 676 EGF-LIKE 17.  
FT DOMAIN 678 714 EGF-LIKE 18.  
FT DOMAIN 716 751 EGF-LIKE 19.  
FT DOMAIN 753 789 EGF-LIKE 20.  
FT DOMAIN 791 827 EGF-LIKE 21.  
FT DOMAIN 829 867 EGF-LIKE 22.  
FT DOMAIN 869 905 EGF-LIKE 23.  
FT DOMAIN 907 943 EGF-LIKE 24.  
FT DOMAIN 945 981 EGF-LIKE 25.  
FT DOMAIN 983 1019 EGF-LIKE 26.  
FT DOMAIN 1021 1057 EGF-LIKE 27.  
FT DOMAIN 1059 1095 EGF-LIKE 28.  
FT DOMAIN 1097 1143 EGF-LIKE 29.  
FT DOMAIN 1145 1181 EGF-LIKE 30.  
FT DOMAIN 1183 1219 EGF-LIKE 31.  
FT DOMAIN 1221 1265 EGF-LIKE 32.  
FT DOMAIN 1267 1305 EGF-LIKE 33.



```
FT DOMAIN 1307 1346 EGF-LIKE 34.
FT DOMAIN 1348 1384 EGF-LIKE 35.
FT DOMAIN 1387 1426 EGF-LIKE 36.
FT REPEAT 1445 1480 LIN/NOTCH 1.
FT REPEAT 1481 1522 LIN/NOTCH 2.
FT REPEAT 1523 1562 LIN/NOTCH 3.
FT REPEAT 1917 1946 ANK 1.
FT REPEAT 1950 1980 ANK 2.
FT REPEAT 1984 2013 ANK 3.
FT REPEAT 2017 2046 ANK 4.
FT REPEAT 2050 2079 ANK 5.
FT DOMAIN 1730 1733 POLY-ALA.
FT DOMAIN 1891 1894 POLY-GLU.
FT DOMAIN 2258 2261 POLY-PRO.
FT DOMAIN 2497 2500 POLY-SER.
FT SITE 1654 1655 CLEAVAGE BY URIN-LIKE PROTEASE (BY
SIMILARITY).
FT DISULFID 24 37 BY SIMILARITY.
FT DISULFID 31 46 BY SIMILARITY.
FT DISULFID 48 57 BY SIMILARITY.
FT DISULFID 63 74 BY SIMILARITY.
FT DISULFID 68 87 BY SIMILARITY.
FT DISULFID 89 98 BY SIMILARITY.
FT DISULFID 106 117 BY SIMILARITY.
FT DISULFID 111 127 BY SIMILARITY.
FT DISULFID 129 138 BY SIMILARITY.
FT DISULFID 144 155 BY SIMILARITY.
FT DISULFID 149 164 BY SIMILARITY.
FT DISULFID 166 175 BY SIMILARITY.
FT DISULFID 182 195 BY SIMILARITY.
FT DISULFID 189 204 BY SIMILARITY.
FT DISULFID 206 215 BY SIMILARITY.
FT DISULFID 222 233 BY SIMILARITY.
FT DISULFID 227 243 BY SIMILARITY.
FT DISULFID 245 254 BY SIMILARITY.
FT DISULFID 261 272 BY SIMILARITY.

Query Match 2.6%; Score 164; DB 1; Length 2531;
Best Local Similarity 19.8%; Pred. No. 0.26;
Matches 109; Conservative 61; Mismatches 171; Indels 210; Gaps 27;

QY 402 GKDWPDFFYSHSGOKVYLTPVPCNACGPGSWIKGYCDKACNNSACDWDGDCSGN 461
DB 1437 GRDIPP-----PQIEACELPEQCEDAGNK-----VCNLQCNHACGWDGDC--- 1479
QY 462 SSGSRYIAGGGTSGVGPQWQ-----FGGINSYSYCNQ--GC----- 499
DB 1480 -----SLNFNDPMKNTQSLQCKWYFSDG--HCDSCQNSAGCLDFGDCQLT 1524
QY 500 -----ANSWLADKFCQACNLVSCGFDAGCGODHPHE-----LYKVILLP-- 540
DB 1525 EGOCNPLYDQYKDHFDHGDHCGCNSACEWDGLDCA-EHVPRLAAGTILVLLVLLPD 1583
QY 541 -----NOTHYIIPKGECL-----PYFSPAFAVAKRGVEGAYSNDPIIRHA 579
DB 1584 QLRNNSFHLRDYSHVLTHTNVFKRDAQOQQMIFPY-----GREELRKHPIKRS 1635
QY 580 SIANKWKTTHLI-----MHSGMATTTHFNLTQNTNDEEFKQITVEVDTRE----- 627
DB 1636 V---GWATTSLLPGTNGGQRRELDPMIDHGSIVY-----LEIDNROCVOSS 1679
QY 628 -----GPKLNSTAGQYENLVSPITLLPEABILLPEDIP----- 660
DB 1680 SQCFQSATDVAFLGALGALSLSINIPYKTEAV-KSETVEPPLPSQLHLMYVAAAFVLLF 1738
QY 661 -----KKR-----FKRHDVNSTRRAQEEVKIPL--VNISLLP-KDAQLS 700
DB 1739 FVCGVLLSKRRKHQHQWLFPEGFKVSEAKKRR-----PLGEDSVGLKPLKNASDG 1793
QY 701 LNTLDLQLEHG--DITLKGYNLSKALLRSFLMNSQHKAKIKQAIIITDETNDLSVAPQEK 758
DB 1794 ALMDNQNEWDGEDLETKKFRFEEPVLPDLDDQDTHRQWTOQHLDAADLRVSAMAPTTP 1853
```

```
QY 759 QVHKSILNSLGVSRLOLTFPAVSVKVNCHDOGNPDLLETARPRVETHQKTIGG 818
DB 1854 Q-----GEVDADCDMDNVNVRGPD--GFTPLMIASCSGGLET-----GN 1889
QY 819 NVTKKEKPPSLI 829
DB 1890 SEEEEDAPAVI 1900

RESULT 8
NTCL_BRARE STANDARD; PRT; 2437 AA.
AC P46530;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Neurogenic locus notch homolog protein 1 precursor.
GN NOTCH1A OR NOTCH.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OC NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA MEDLINE=94128602; PubMed=8297791;
RX Bierkamp C.; Campos-Ortega J.A.;
RT "A zebrafish homologue of the Drosophila neurogenic gene Notch and
RL Mech. Dev. 43:87-100(1993).
CC -!- FUNCTION: IMPLICATED IN CELL FATE SPECIFICATIONS DURING OF THE
CC EMBRYO DEVELOPMENT. MAY BE INVOLVED IN THE FORMATION OF THE
CC NEURAL PLATE, NOTOCHORD AND BRAIN VESICLES.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- DEVELOPMENTAL STAGE: Expressed in all cells in pregastrulation
CC stages. During gastrulation is differentially expressed,
CC accumulating predominantly in the prechordal mesoderm and
CC notochord. At the end of gastrulation, expressed along the
CC anterior-posterior axis including the developing neural plate
CC and differentiating mesoderm. Also present in the developing
CC brain and head regions.
CC -!- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
CC -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X69088; CAA48831.1; -
CC HSPSP; P00740; IEDM.
CC ZFIN; ZDB-GENE-990415-173; notchla.
CC InterPro; IPR002110; ANK.
CC InterPro; IPR000152; Asx_hydroxyl.
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR000742; EGF-2.
CC InterPro; IPR001881; EGF-CA.
CC InterPro; IPR001438; EGF-II.
CC InterPro; IPR000800; Notch.
CC Pfam; PF00008; EGF; 36.
CC Pfam; PF00023; ank; 6.
CC Pfam; PF00066; notch; 3.
CC PRINTS; PR00010; EGFBL00D.
CC PRINTS; PR01452; NOTCH.
CC SMART; SM00248; ANK; 5.
CC SMART; SM00179; EGF_CA; 19.
CC SMART; SM00001; EGF_like; 16.
```





Db 2130 K-----NEVSIKALQVYGNRIFTDKNGKGGISFFIPPIKODTLKTIINETI 2181  
Qy 811 -----HTOKTIGNVTK-----EKPPSILVPLESOMTKKTKCKENSRME 853  
Db 2182 DNSNIKQRLGIYIFVRKNVSENSKLCDFTTGTSLSMELNSQV-KEKCKTVKIKRGDIFG 2240  
Qy 854 ENAENHIGV-----TEVLLGRKLQHYTDS-YLGFPLWEKKY-----FLLDDEESL 900  
Db 2241 LKCPKGAIFPQACSFNSVLLVEYKSDYEDSEHINYIHKDKKYNLKPDKVIELMDEN--- 2297  
Qy 901 KQLAYFTDSKNRKYR-DTFASLRVYN-KILNSKFGFTSRKVPAAH--MPHMIDRIVM 956  
Db 2298 -----FRELQIQOYTGISNITDVLHPKFNGLNPLNFKNHYSTAYAKVPDTFNSIIN 2351  
Qy 957 QELQDMRPE-----EFDKTSFKVHRHSEDMOFAP-----SYFYLLMSAVOPLNIS 1001  
Db 2352 FSCNCPNPKHYVTQMQUESNRPNFNKKNWIKNELLNPKIEYALLDDEERQKKIK 2411  
Qy 1002 QVFDEVTDSQSVLSDRIRTLATRIHELPLSLDGLTGLEHMLI----- 1045  
Db 2412 QQEEEQEQE-ILKQDDR-----LSRHDDYKNKHTYILYDSNEHICDYKNES 2459  
Qy 1046 -----NCSKMLPADITQLN-----NIPPTQESYYDNLPPVTSKLVNLC 1084  
Db 2460 LISTLPNDTKKIQKSICKINAKALDVVTKPHTKNETPKD---YFPN-----SSLITND 2511  
Qy 1085 KPVTDKIHK-----AYKDKNRYFEI 1105  
Db 2512 KKIIVTFDKNFVYIYIDTKTFFSL 2536

RESULT 10  
RA50\_METJA STANDARD; PRT; 1005 AA.  
ID RA50\_METJA  
AC Q58718;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE DNA double-strand break repair rad50 AtPase.  
GN RAD50 OR MJ1322.  
OS Methanococcus jannaschii.  
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
OC Methanocaldococcaceae; Methanocaldococcus.  
OX NCBI\_TaxID=2190;  
RN [1]  
RC SEQUENCE FROM N.A.  
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;  
RX MEDLINE=96337993; PubMed=8688087;  
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glode A.,  
RA Scott J.L., Geohagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
jannaschii";  
RL Science 273:1058-1073(1996).  
CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The  
CC rad50/mrell complex possesses single-strand endonuclease activity  
CC and ATP-dependent double-strand-specific exonuclease activity.  
CC Rad50 provides an ATP-dependent control of mrell by unwinding  
CC and/or repositioning DNA ends into the mrell active site (By  
CC similarity).  
CC -!- SUBUNIT: Forms a complex with mrell (By similarity).  
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC DR EMBL; U67572; AAB99331.1; -.  
DR TIGR; MJ1322; -.  
DR InterPro; IPR003439; ABC\_transportr.  
DR InterPro; IPR003405; SMC\_C.  
DR InterPro; IPR003395; SMC\_N.  
DR Pfam; PF02463; SMC\_N; 1.  
DR Pfam; PF02483; SMC\_C; 1.  
DR ProDom; PD000006; ABC\_transportr; 1.  
KW DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.  
FT NP\_BIND 32 39 ATP (BY SIMILARITY).  
FT DOMAIN 158 849 COILED COIL (POTENTIAL).  
SQ SEQUENCE 1005 AA; 119387 MW; 9BBB48173E78F3 CRC64;  
  
Query Match 2.6%; Score 162.5; DB 1; Length 1005;  
Best Local Similarity 19.2%; Pred. No. 0.081;  
Matches 139; Conservative 124; Mismatches 255; Indels 205; Gaps 33;  
  
Qy 593 NKWKTILHLSGNNATTIHEN-LTFQNTDEEFKMQITVEVDTRREGPKLNSTAQKGYEN 641  
Db 232 NEWENKKLLYEKFKLEERKRALELKNQELKILEYDUNTVVEARETUNRHKDEYEKYKS 291  
Qy 642 LVSEPT-----TLLPEAEILFEDIPKEKRF-PKFK-RHDVNSTRAQE 681  
Db 292 LVDEIRKTESRLRELKSHVEDYLLTKOLEIKGDIKLEKFKINFSKYRDDIDNLDLTLN 351  
Qy 682 EVKIPLVNISLLPKDAQLSLNTLDLQLEHGDITLKGYNLSKALLRSFLMSQHAQIKNQ 741  
Db 352 KIKDEIERVETI-KDLLEELKNLEEEI-----KIEKYKRICECKEYK---EKYLEBEK 403  
Qy 742 AIIITDE-TNDSLVAPOEKOVHKSILPNSLGVSELRORLTFPAVSVKVNGHGOQNPPDL 800  
Db 404 AVEYNKLTLEVITLLOEK---KSTEKINDLETRINKLL-----BETKN--IDI 447  
Qy 801 ETTARFRVETHQKTIGONVTKEKPP--SLIVPLESQMTEKKITGKEK----- 847  
Db 448 ESIENSLKEIEBKRVLENLQKEKTELKNGEINSEIKRLKILDELKEVEGRCPLCKT 507  
Qy 848 ---ENSRMEENAHNIGVTEVLLGRKLQHYTDSYLGFLPWEKKYFLLD-----LDE 896  
Db 508 PIDENKME-----LINQHTQLANKYTELEBEINKKIREIEKDEKLKKEIDK 555  
Qy 897 EESLKTQLAYFTDSKNR---ARYKRDTPADSLRYVYNK-----ILNSKFGFTSRKVPAAHP 948  
Db 556 EENLTKLTLYLEKQSQIELELELKNYKEQLDEINKKISNVYNGK----- 602  
Qy 949 HMIDRVNQELQDMFPEEDFKTSFHKVHRHSEDMOFAFSYFYLLMSAVOPLN-----ISQ 1002  
Db 603 -PVDEI-----LADI-----KSQLNKEK-----NFYNYLSAVSYLSNSVDEGIRN 642  
Qy 1003 VFDEVDVTDQSG-----VLSDEIRTLATRIHELPLSLDGLTGLEH-MLNCS 1048  
Db 643 RIKEITENIVGWNKEKREELNKLUREDERENLRUKLNEKKEKLEIELENRSKLPD 702  
Qy 1049 KW-----LPADITQILNIPPTQESYDNPPLPVTKSLVNTCKPVT-DKIHKAYKDKNKR 1102  
Db 703 KYKEVLGLTEKLEELKNKIDGLEIYN-----ICNSKILADINKRY---NKED 749  
Qy 1103 FEIWEDEEI-----AFKMRITNVSHVQGLDDIRKNP---RKVCLNDN----- 1143  
Db 750 IEIVLNNKILEVNKEINDIEERISVINGKLDIEINYNEEHKKIKELYENKQELDNVREQ 809  
Qy 1144 -----LDHNNKDAQTVKAVLROFYESMFFIPSOQELPREYRNRFLLHHELQEWRAVD 1196  
Db 810 KTEIETGLEYLKDKOVESLARKEM-----SMLEKEKEKLTKEVE-----YID 852  
  
Qy 1197 KIK 1199  
Db 853 KVR 855

RESULT 11  
NTCL\_MOUSE  
ID NTCL\_MOUSE STANDARD; PRT; 2531 AA.  
AC 001705; Q90W58; Q99JC2; Q06007; Q61905; Q9R0X7;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Neurogenic locus notch homolog protein 1 precursor (Notch 1) (Notch A) (MRL4) (p300).  
GN NOTCH1 OR NOTCH.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RC TISSUE=Embryo;  
RX MEDLINE=93194170; PubMed=8449489;  
RA Franco del Amo F., Gendron-Maguire M., Swiatek P.J., Jenkins N.A.,  
RA Copeland N.G., Gridley T.;  
RT "Cloning, analysis, and chromosomal localization of Notch-1, a mouse  
RT homolog of Drosophila Notch.";  
RL Genomics 15:259-264(1993).  
RN [2]  
RP SEQUENCE OF 731-1899 FROM N.A. (ISOFORM 2), AND DEVELOPMENTAL STAGE.  
RC STRAIN=CD-1; TISSUE=Embryo;  
RX MEDLINE=93050801; PubMed=1426644;  
RA Reaume A.G., Conlon R.A., Ziringibl R., Yamaguchi T.P., Rossant J.;  
RT "Expression analysis of a Notch homologue in the mouse embryo.";  
RL Dev. Biol. 154:377-387(1992).  
RN [3]  
RP SEQUENCE OF 1551-1647 FROM N.A. (ISOFORM 1), AND DEVELOPMENTAL STAGE.  
RC TISSUE=Embryo;  
RX MEDLINE=93048835; PubMed=1425352;  
RA Greenspan R.J., McMahon A.P., Gridley T.;  
RT "Expression pattern of Notch, a mouse homolog of Drosophila Notch,  
RT suggests an important role in early postimplantation mouse  
RT development.";  
RL Development 115:737-744(1992).  
RN [4]  
RP SEQUENCE OF 1161-1547 FROM N.A.  
RC STRAIN=C57BL/6 X CBA; TISSUE=Embryo;  
RX MEDLINE=93178563; PubMed=8440332;  
RA Lardelli M., Lendahl U.;  
RT "Notch A and Notch B-two mouse Notch homologues coexpressed in a  
RT wide variety of tissues.";  
RL Exp. Cell Res. 204:364-372(1993).  
RN [5]  
RP SEQUENCE OF 1659-1673 FROM N.A.  
RX MEDLINE=99364499; PubMed=10437788;  
RA Lee J.S., Ishimoto A., Yanagawa S.I.;  
RT "Murine leukemia provirus-mediated activation of the Notch1 gene leads  
RT to induction of HES-1 in a mouse T lymphoma cell line, DL-3.";  
RL PNAS Lett. 455:276-280(1999).  
RN [6]  
RP SEQUENCE OF 1950-2201 FROM N.A.  
RX MEDLINE=98029496; PubMed=9384671;  
RA Messerle M., Follo M., Nehls M., Eggert H., Boehm T.;  
RT "Dynamic changes in gene expression during in vitro differentiation of  
RT mouse embryonic stem cells.";  
RL Cytokines Cell. Mol. Ther. 1:139-143(1995).  
RN [7]  
RP SEQUENCE OF 1655-1659, CLEAVAGE BY FURIN-LIKE CONVERTASE, AND  
RP MUTAGENESIS OF 1651-ARG-GLN-ARG-ARG-1654.  
RX MEDLINE=98318619; PubMed=9653148;  
RA Logeat F., Bessia C., Brou C., Leblat O., Jarriault S., Seidah N.G.,  
RA Israel A.;  
RT "The Notch1 receptor is cleaved constitutively by a furin-like  
RT convertase.";  
RL Proc. Natl. Acad. Sci. U.S.A. 95:8108-8112(1998).  
RN [8]

PARTIAL SEQUENCE, AND POST-TRANSLATIONAL PROCESSING.  
MEDLINE=21523956; PubMed=11518718;  
Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;  
RT "Murine notch homologs (NL-4) undergo presenilin-dependent  
RT proteolysis.";  
RJ Biol. Chem. 276:40268-40273(2001).  
RN [9]  
RP POST-TRANSLATIONAL PROCESSING.  
RX MEDLINE=21374376; PubMed=11459941;  
RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;  
RT "Conservation of the biochemical mechanisms of signal transduction  
RT among mammalian Notch family members.";  
RJ Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).  
CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands  
CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.  
CC Upon ligand activation through the released notch intracellular  
CC domain (NICD) it forms a transcriptional activator complex with  
CC RBP-J kappa and activates genes of the enhancer of split locus.  
CC Affects the implementation of differentiation, proliferation and  
CC apoptotic programs (By similarity). May play an essential role in  
CC postimplantation development, probably in some aspect of cell  
CC specification and/or differentiation. May be involved in mesoderm  
CC development, somite formation and neurogenesis. Involved in the  
CC maturation of both CD4+ and CD8+ cells in the thymus.  
CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-  
CC terminal fragment N(EC) which are probably linked by disulfide  
CC bonds.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following  
CC proteolytical processing NICD is translocated to the nucleus.  
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be  
CC produced by alternative splicing.  
CC -1- TISSUE SPECIFICITY: Highly expressed in the brain, lung and  
CC thymus. Expressed at lower levels in the spleen, bone-marrow,  
CC spinal cord, eyes, mammary gland, liver, intestine, skeletal  
CC muscle, kidney and heart.  
CC -1- DEVELOPMENTAL STAGE: First detected in the mesoderm at 7.5 d.p.c.  
CC By 8.5 d.p.c. highly expressed in presomitic mesoderm, mesenchyme  
CC and endothelial cells, while much lower levels are seen in the  
CC neuroepithelium. Between 9.5-10.5 d.p.c. expressed at high levels  
CC in the neuroepithelium. At 13.5 d.p.c. expressed in the surface  
CC ectoderm, eye and developing whisker follicles.  
CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form  
CC which is proteolytically cleaved by a furin-like convertase in the  
CC trans-Golgi network before it reaches the plasma membrane to yield  
CC an active, ligand-accessible form. Cleavage results in a C-  
CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following  
CC ligand binding, it is cleaved by TNF-alpha converting enzyme  
CC (FACE) to yield a membrane-associated intermediate fragment called  
CC notch extracellular truncation (NEXT). This fragment is then  
CC cleaved by presenilin dependent gamma-secretase to release a  
CC notch-derived peptide containing the intracellular domain (NICD)  
CC from the membrane.  
CC -1- PTM: Phosphorylated.  
CC -1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.  
CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.  
CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.  
CC -1- SIMILARITY: CONTAINS 5 ANK REPEATS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Z11886; CAA77941.1; -;  
DR EMBL; L02613; XAK14898.1; -;  
DR EMBL; X68278; CAA48339.1; -;  
DR EMBL; AJ238029; CAB40733.1; -;  
DR EMBL; X82562; CAA57909.1; -;  
DR HSSP; P00740; IEDM.  
DR MGI; 97363; Notch1.

QY	555	YFSPAENVAKGVGAGVSDNPPIIRHSIAK-----WKTHILI-----MHSGMNA	598
Df	1599	VLHTNVVKRDAQGQQMIFFYYGHBEELRKHPIKRSTVGWATSSLPGTSGGRQRRELDIP	1658
QY	599	TTTHENLTFONTNDEEFKMGIIVEDVTRE-----GPKLNSTAQKGYNILV	643
Df	1659	MDRGSIVY-----LEIDNRQVOSSQCFSQSFATDVAAPFALGALSIGLSNIIP	1705
QY	644	SPITLLPEAEIIFEDIP-----KEKR-----FKP-FK	669
Df	1706	YKIEAV-KSEPVPEPLPSQLHLMYAAAAFVLFFWGCGVILLSRRRHQHGLHFPFGFK	1764
QY	670	RHDVNSTRRAQEVEKIPL--VNISLLP-KDAOLSUNTLDLGLEHG--DIITLGKYNLSKSA	724
Df	1765	VSRASKKKKRE-----PLGEDSVGLPKPLKNASDGALMDNDNQMGWDEDLETKKFRFEPPV	1819
QY	725	LLRSFLMSHQAHAIKNOAIIITDETNSLVAPQEKOVHKSIPLNSLIGVSVERLORLTFFPAVS	784
Df	1820	VFDPDSQTDHRQTQOHHDAADLRMSAWPTTPQ-----GEVDADCMD	1863
QY	785	KVYNGHDGOQNPLLDTETTRARFVRTHTQKTIGGNVTKEKPSPSLI-----VPLESQMT	837
Df	1864	VNVRGPD--GFPTPMIASCSGGGLT-----GNSEEREADAPAVISDFIYGASLNQTD	1915
QY	838	KEKKIIT-----GREKENSRMEENAEVHI-----GVTEVILL	867
Df	1916	RTGETALTALAARYSRRDRKRILEASADANIQDNMGRTPLHAHVASADAQGQVFQIILL	1970
RESULT 12			
MYH9_RAT			
ID	MYH9_RAT	STANDARD; PRT; 1961 AA.	
AC	G62812;		
CD	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DD	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Myosin heavy chain, nonmuscle type A (Cellular myosin heavy chain,		
DEF	type A) (Nonmuscle myosin heavy chain-A) (NMHC-A).		
GN	MYH9.		
OS	Rattus norvegicus (Rat).		
OX	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Musinae; Rattus.		
NCBI_TaxID=10116;			
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Choi O.H., Park C., Itoh K., Adelstein R.S., Beaven M.A.;		
RL	Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.		
CC	-!- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS,		
CC	CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND		
CC	CAPPING.		
CC	-!- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY		
CC	CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2		
CC	REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).		
CC	-!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE. SHOWING		
CC	CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,		
CC	CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.		
CC	-!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.		
CC	-!- SIMILARITY: CONTAINS 1 IQ DOMAIN.		
-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collabora-		
CC	tion between the Swiss Institute of Bioinformatics and the EMBL Outstatid-		
CC	the European Bioinformatics Institute. There are no restrictions on		
CC	use by non-profit institutions as long as its content is in no		
CC	modified and this statement is not removed. Usage by and for commerc-		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announ-		
CC	or send an email to license@isb-sib.ch).		
-----			
EMBL:	U31463; AAA74950.1; --		
DR	HSPG; PI0587; IBR2.		
DR	InterPro: IPRO00048; IQ_region.		
DR	InterPro: IPRO004009; Myosin_N.		
DR	InterPro: IPRO02928; Myosin_tail		



```
DR InterPro: IPR001609; myosin_head.
DR Pfam: PF00063; myosin_head; 1.
DR Pfam: PF00612; IQ; 1.
DR Pfam: PF01576; Myosin_tail; 1.
DR Pfam: PF02736; Myosin_N; 1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR ProDom: PD000355; myosin_head; 1.
DR SMART: SM00015; IQ; 1.
DR SMART: SM00242; MYSC; 1.
DR PROSITE: PS00096; IQ; 1.
DR Myosin: ATP-binding; Calmodulin-binding; Actin-binding;
KW Coiled coil; Alkylation; Multigene family.
FT DOMAIN 1 778 MYOSIN HEAD-LIKE.
FT DOMAIN 779 808 IQ.
FT DOMAIN 841 1927 COILED COIL (POTENTIAL).
FT NP_BIND 174 181 ATP (POTENTIAL).
FT DOMAIN 654 676 ACTIN-BINDING.
FT MOD_RES 694 694 ALKYLATION (SH-1) (POTENTIAL).
FT MOD_RES 704 704 ALKYLATION (SH-2) (POTENTIAL).
FT SEQUENCE 1961 AA; 226336 MW; 99876D9681F819E CRC64;

Query Match 2.5%; Score 158.5; DB 1; Length 1961;
Best Local Similarity 21.6%; Pred. No. 0.38;
Matches 121; Conservative 104; Mismatches 246; Indels 89; Gaps 23;

QY 622 EVDTRGPKLNSTAGQYENLVSPITLLPEAEILLFEDIPKRPKFRKHDVNSTRRQAE 681
DB 870 EMTWOSQLMAEKLOL-QEQQAQTELCAEAEELRATKQOELEETCHDLEA-RVEEE 927
QY 682 EYKIPLVNISLLPKDAQSLNLDLQLEHGDTTLKGYNLK---SALLR-----SFLMNS 733
DB 928 EERCQYLQAE--KKMQQNIQLEQLBEESARQKQLEKVTYAKLKLLEEQIIMED 985
QY 734 QHAKI-KNOAITD---ETNDSLVAPOEQVHKHSILPNS-----LGVSERLQRTFFPAVS 784
DB 986 QNCKLAKEKLEEDRAVEFTTDLMEEBEKSLSLAKLNKHEAMITDLERLRR----- 1038
QY 785 VKVNGHDQGNPDLDTTARFVETHQKTIGGNVTKKPPSLVPLESQMTEKKITG 844
DB 1039 -----EEKQOELE-----KTRKLEGDSTDLN--DQTAELQAOIAELKMQLA 1079
QY 845 KEKEN-----SRMEEN-AENHIGVTEVLGKRLQHYTDSYLGLPWKEK-----KKYFLD 892
DB 1080 KKEEELQALAREVEEAQKNALKI---RELETQISELOEDLESEACRNKAQKQRD 1136
QY 893 LIDEESLTKQLAYTFDSKNRARYKRDFTADSLRVYVKNILNSKFGFTSRKVPAPHMID 952
DB 1137 LGEELEALKTELEDTLIDSTAAQOELRSKREQEVSLKKTLEDE---AKTHEAQIQEM-- 1190
QY 953 RIVMQELODMFPPEFDKTSFHKVRHSDMQAFSYFYILMSAVOPL-----NISQV 1003
DB 1191 ROKHSQAVELEAQLEQTKRVKATLEKAKQTLNENGERLANEVKALLOCKGDSEHKRVK 1250
QY 1004 FDEVDTQSGVLSDRIRI-TLATRIHELPLSLQDLTGLFHLMLINGSKMLPADITOL--NNI 1061
DB 1251 EAQLQELQVKFSEGERVTRADKYSKQLQVELDSVTGLLNQSDSKSKLTKDFSALSQL 1310
QY 1062 PPTQESYDNPDPVTKSLVTNCKPVTDKIHKAYDKKNKYRFEIMGEBEIEAFKMTRTNVS 1121
DB 1311 QDTQELLEENRQKL--SLSTKLQMED-----EKNSFRQLEBEEBEEAKRNLEKQTA 1361
QY 1122 HVVGQLDDIRKPNKRV-CL 1140
DB 1362 TLHAQVTDMMKKMEDGVGCL 1381

RESULT 13
ID SBCC_CLOAB STANDARD; PRT; 1163 AA.
AC Q97FK1;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Nuclease sbcCD subunit C.
GN SBCC OR CAC2736.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=2159325; PubMed=11466286;
RA Noelling J., Breton G., Omeichenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RA "Genome sequence and comparative analysis of the solvent-producing
bacterium Clostridium acetobutylicum."
RT J. Bacteriol. 183:4823-4838(2001).
RL -!- FUNCTION: SbcCD cleaves DNA hairpin structures. These structures
CC can inhibit DNA replication and are intermediates in certain DNA
CC recombination reactions. The complex acts as a 3'->5' double
CC strand exonuclease that can open hairpins. It also has a 5'
CC single-strand endonuclease activity (by similarity).
CC -!- SUBUNIT: Heterodimer of sbcC and sbcD (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. SBCC SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE007771; AK80682.1; -.
CC InterPro: IPR003439; ABC_transport.
CC Hydrolase; Nuclease; Exonuclease; Endonuclease; DNA replication;
CC DNA recombination; ATP-binding; Coiled coil; Complete proteome.
CC NP_BIND 35 42 ATP (POTENTIAL).
CC DOMAIN 197 415 COILED COIL (POTENTIAL).
CC FT DOMAIN 446 1003 COILED COIL (POTENTIAL).
CC FT SEQUENCE 1163 AA; 135507 MW; CE5F0BD2215D7A92 CRC64;

Query Match 2.5%; Score 158; DB 1; Length 1163;
Best Local Similarity 21.5%; Pred. No. 0.19;
Matches 129; Conservative 95; Mismatches 195; Indels 182; Gaps 31;

QY 630 KLNSTAQK-----YENLVSPITLLPEAEILLFEDIPKRPKFRKHDVNSTRRQEEV 683
DB 501 KLNDSRQKWAKYSEYNSLAKSLRVENSEQVLR--TKKEMTKL----- 545
QY 684 KIPLVNTSLLPKDAQSLNLDLQLEHG-----DITLKGYNLSKALL 726
DB 546 KISKVNKIESLETENNAHVLRKLSGECAPCGSVHHKEGFKVDLKAETLKSE-L 604
QY 727 RSFLMNSQAHKKNQAIITDETNDLSLVAPOEQVHKHSILPNSLG-----YSERLQRTFF 781
DB 605 EGF---EKRRFEEIEMCEASIKVEKNIKKLNESI--NNLGEFKEVLESMEKKFN 659
QY 782 AVSVKVGND-----QGNPPDL-L-ETTARFVETHQKTIGGNVTKKPPSLVPLESQM 836
DB 660 YLKEKVNKFNLEKIQDNDNKDLSERSNKIEVEYQKEKTKVQCEKR-----IVDLKSEL 714
QY 837 TK-----EKKITG-----KEKNSRMEENAEHNHIGVTEVLGRKL 871
DB 715 EEAKEFNEVAYTIENKAEKIQDFEMKEILEKERVVE--AEGEIKDLNLL--NI 770
QY 872 QHYTDSYGLFWPEKKYFFDLLDEESLTKQLAYFTDSKNRARYKRDFTADSLRVYVKNKI 931
DB 771 RH-----TEKEQ---LMDKCSRLKEEL-----SKNKAELKEK-----DKI 802
QY 932 LNSKFGFTSRKVPK-----HMPHMDIRV-----MOELODMPEEFDKTSFH 973
DB 803 INEKIELKKNVGVLDNLNLYELKEIKETGKIEBQYNLCOKKMNIEED----KYRCSDE 858
```



```

QY 974 KVRHSEDMQAFYFYLYMSAVQPLN---ISQVEVDVTDQSGVLSREINTLATRIHEL 1030
DQ 859 IIKVHNSLS---SLKDRKVDNDKLNKILMEKEFENIEKAKENYLNKLSQVYKY 915
QY 1031 PLSLODLTGLEHMLNSKMLPADITQNLNPPPOESYYPDNLPPVTKYSLVNTCKPVTDK 1090
DQ 916 KNEUSKVGAVEVL---SK-----KLKNRKLTEEKWIE-----IQN-----NR 950
QY 1091 IHKAYKDKNKYRFEIMGEETI-----AFKMIRT--NVSHVVGOLDIDRK--NPRKF 1137
DQ 951 VEKASKAKALQERSIKLEEEVKNTEIKELKGLKLYKQELHKLKLLDLEKLFKQKGF 1010
QY 1138 V 1138
DQ 1011 V 1011

RESULT 14
MYHB_CHICK
ID MYHB_CHICK STANDARD; PRT; 1978 AA.
AC P10587;
DT 01-JUL-1989 (Rel. 11, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, gizzard smooth muscle.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=8611818; PubMed=2892941;
RA Yanagisawa M., Hamada Y., Katsuragawa Y., Inamura M., Mikawa T.,
RA Masaki T.;
RT "Complete primary structure of vertebrate smooth muscle myosin heavy
RT chain deduced from its complementary DNA sequence. Implications on
RT topography and function of myosin.";
RL J. Mol. Biol. 198;143-157(1987).
RN [2]
RP REVISIONS.
RA Masaki T.;
RL Submitted (FEB-1989) to the EMBL/GenBank/DBJ databases.
RN [3]
SEQUENCE OF 1-203.
RX MEDLINE=88032919; PubMed=3312184;
RA Maizumi T., Onishi H., Yajima E., Matsuda G.;
RT "Amino acid sequence of the amino-terminal 24 kDa fragment of the
RT heavy chain of chicken gizzard myosin.";
RL J. Biochem. 102;133-145(1987).
RN [4]
X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS) OF 1-818, AND REVISIONS 204-215.
RX MEDLINE=98412652; PubMed=9741621;
RA Dominguez R., Freyza Y., Trybus K.M., Cohen C.;
RT "Crystal structure of a vertebrate smooth muscle myosin motor domain
RT and its complex with the essential light chain: visualization of the
RT pre-power stroke state.";
RL Cell 94:559-571(1998).
CC -1- FUNCTION: MUSCLE CONTRACTION.
CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMW). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).

```

```

CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X06546; CAA29793.1; -.
DR PIR; S03166; S03166.
DR PDB; 1BR1; 09-SEP-98.
DR PDB; 1BR2; 09-SEP-98.
DR PDB; 1BR4; 09-SEP-98.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
DR InterPro; IPR002017; Spectrin.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50096; IQ; 1.
DR Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Methylation; Alkylation; Calmodulin-binding;
FT MULTIGENE FAMILY; 3D-structure.
FT INIT_MET 0
FT DOMAIN 1 790 MYOSIN HEAD-LIKE.
FT DOMAIN 791 820 IQ.
FT DOMAIN 849 1978 RODLIKE TAIL (S2 AND LMM DOMAINS).
FT DOMAIN 849 1978 COILED COIL (POTENTIAL).
FT NP_BIND 176 183 ATP.
FT DOMAIN 566 588 ACTIN-BINDING.
FT DOMAIN 767 781 ACTIN-BINDING.
FT MOD_RES 1 1 BLOCKED.
FT MOD_RES 127 127 METHYLATION (TRI-) (POTENTIAL).
FT MOD_RES 706 706 ALKYLATION (SH-1).
FT MOD_RES 716 716 ALKYLATION (SH-2).
FT CONFLICT 127 127 MISSING (IN REF. 3).
FT CONFLICT 204 215 KDTISITQGPSFS -> RTPASLVKVLPP (IN REF. 1).
SQ SEQUENCE 1978 AA; 228663 MW; B7B6C923E5273D93 CRC64;

Query Match 2.5%; Score 157.5; DB 1; Length 1978;
Best Local Similarity 20.8%; Pred. No. 0.44;
Matches 126; Conservative 122; Mismatches 250; Indels 109; Gaps 28;

QY 612 DEEFKMQITVEVDTRFGPKLNSTAGK-----YENLVSPITLLPEAEILFEDIPK 661
DQ 863 DEE---LQRTKERQQAEEAELEKELQKLTQCEKNLQKLAETELYAAEENRVLAA 920
QY 662 EKRPFKRHDVNSTRRQAQEEVKIPLVNISSLPKDAQLSLNTLDLQLEHGDITLKGYNLS 721
DQ 921 KQOELEELHEWEARIEEEERS---QOLQAEKKKMOOQMLDLEEQEEEAARQKLOLE 977
QY 722 KSALLRSFLMNSHAKIKNQAITDETNDLSVAPQ---EKQVHKSILPNSLGVSRQLR 778
DQ 978 K-----VTADGKIKKMWEDDILIMEDQNNKLTKEKLLSEEV-SDLTNLAAEEAKNL 1030
QY 779 TFPADVSVKNGHDGQNPDL-----ETTARFRVETHTQKTIGCNVTKEKPPSLIVPLES 834
DQ 1031 T-----KLKNKHEMISELEVRLLKKEKSRQELKIRKLEGESSDLHEQ---IAELQA 1081
QY 835 QMTKEKKTGKEKN-----SRME-ENAEHNHIGVTEVLLGRKLQHYDYSIGLFWPEK-- 886
DQ 1082 QIAELKQAQAKKEELQAAALAEDETSQNNALKKI---RELESHISDLEQLESEKAA 1138

```

QY 887 -----KKYFLDLLDEESLKTQLAYFTD---SKNRYKRDFTADSLRYVKNILNSKFGFT 939  
 Db 1139 RNKAQKQKRLSELEAKTELEDTLDTATQELRAKRE--QEVTLKRALEEE---1191  
 QY 940 SRKVPAMPHWIDR--IVMOELODMPEEF-----DKTSFKHVSHSEDMQFAFSYFF 990  
 Db 1192 TRTHAQVQMRQKHTQAVEELTEQL-EQFKRAKALDKTKQLEKDNADL-----1241  
 QY 991 LMSAVQPLNTSQVFEVD-----TDOSGVLSDR-IRT-LATRIHELPLSLQDLT 1038  
 Db 1242 ---ANEIRLSQAKQDVEHKKKLEVLQQLQSKYSDGERVTELNKVKHKLQIEVENT 1298  
 QY 1039 GLEHMLNCSKMLPADITQL-NNIPPTQESYDNPVPTKSLVTNCKPVTDKIKHAYKD 1097  
 Db 1299 SLLNEASKNKLTQVATLGSQLODTQEL-----LQETROKLN-----VTTKLRLQLED 1349  
 QY 1098 KNYRFFELMEGEETAFKMITNTVSHVVGQLDDIRKPNRFVCLNDINHDHDKAQ-TVKA 1156  
 Db 1350 KNSLQ-BOLDEVEEAKONLERHISTLTQLSDSKKLLQEFATVETMEEGKKLQREIES 1408  
 QY 1157 VLRFYE 1163  
 Db 1409 LTQOFEE 1415

RESULT 15  
 NTC3\_HUMAN STANDARD: PRT; 2321 AA.  
 AC Q9UM47; Q9Y6L8; Q9UPL3; Q9UEB3;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Neurogenic locus notch homolog protein 3 precursor (Notch 3).  
 GN NOTCH3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 EX MEDLINE=97032728; PubMed=8878478;  
 RA Joutel A., Corpechot C., Ducros A., Vahedi K., Chabriat H., Mouton P.,  
 RA Alamowitch S., Domenga V., Cecillon M., Marechal E., Maciazek J.,  
 RA Vayssiere C., Cruaud C., Cabanis E.-A., Ruchoux M.M., Weissbach J.,  
 RA Bach J.-F., Bousser M.-G., Tournier-Lasserre E.;  
 RT "Notch3 mutations in CADASIL, a hereditary adult-onset condition  
 causing stroke and dementia.";  
 RL Nature 383:707-710(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Gunel M., Artavanis-Tsakonas S.;  
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Lamerdin J.E., McCready P.M., Skowronski E., Adamson A.W.,  
 RA Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stillwagen S.,  
 RA Phan H., Velasco N., Ganes J., Danganan L., Poundstone P.,  
 RA Christensen M., Georgescu A., Avila J., Liu S., Attix C., Andreise T.,  
 RA Frankheim M., Amico-Keller G., Coefield J., Duarte S., Lucas S.,  
 RA Bruce R., Thomas P., Quan G., Kronmiller B., Arellano A.,  
 RA Montgomery M., Ow D., Nolan M., Trong S., Kobayashi A., Olsen A.S.,  
 RA Carrano A.V.;  
 RT "Sequence analysis of an 1.5 Mb olfactory receptor (OLFR) cluster in  
 19p13.1";  
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP VARIANTS CADASIL Y-49; C-71; C-90; C-110; C-133; C-141; R-146; C-153;  
 RP C-169; C-171; C-182; R-185; S-212; G-222; Y-224; C-258; Y-542; C-558;  
 RP C-578; C-728; C-985; C-1006; C-1031; C-1231 AND R-1261, AND VARIANTS  
 RP R-170; L-496; Q-1133; M-1183 AND A-2223.  
 RX MEDLINE=98049753; PubMed=9388399;  
 RA Joutel A., Vahedi K., Corpechot C., Troesch A., Chabriat H.,  
 RA Vayssiere C., Cruaud C., Maciazek J., Weissbach J., Bousser M.-G.,

Bach J.-F., Tournier-Lasserre E.;  
 "Strong clustering and stereotyped nature of Notch3 mutations in  
 CADASIL patients.";  
 Lancet 350:1511-1515(1997).  
 RL [5]  
 RN VARIANT CADASIL 114-GLY--PRO-120 DEL.  
 RP MEDLINE=20264473; PubMed=10802807;  
 RX Joutel A., Chabriat H., Vahedi K., Domenga V., Vayssiere C.,  
 RA Ruchoux M.M., Lucas C., Leys D., Bousser M.-G., Tournier-Lasserre E.;  
 RT "Splice site mutation causing a seven amino acid Notch3 in-frame  
 deletion in CADASIL.";  
 Neurology 54:1874-1875(2000).  
 RL [6]  
 RN IDENTIFICATION OF LIGANDS.  
 RP MEDLINE=99180765; PubMed=10079256;  
 RX Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangiu M.-L.,  
 RA Banks A., Lelman J., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;  
 RT "Human ligands of the Notch receptor.";  
 Am. J. Pathol. 154:785-794(1999).  
 CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands  
 Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.  
 Upon ligand activation through the released notch intracellular  
 domain (NICD) it forms a transcriptional activator complex with  
 RBP-J kappa and activates genes of the enhancer of split locus.  
 Affects the implementation of differentiation, proliferation and  
 apoptotic programs (By similarity).  
 CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-  
 terminal fragment N(EC) which are probably linked by disulfide  
 bonds (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following  
 proteolytical processing NICD is translocated to the nucleus.  
 CC -!- TISSUE SPECIFICITY: Ubiquitously expressed in fetal and adult  
 tissues.  
 CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form  
 which is proteolytically cleaved by a furin-like convertase in the  
 trans-Golgi network before it reaches the plasma membrane to yield  
 an active, ligand-accessible form. Cleavage results in a C-  
 terminal fragment N(TM) and a N-terminal fragment N(EC). Following  
 ligand binding, it is cleaved by TNF-alpha converting enzyme  
 (TACE) to yield a membrane-associated intermediate fragment called  
 notch extracellular truncation (NEXT). This fragment is then  
 cleaved by presenilin dependent gamma-secretase to release a  
 notch-derived peptide containing the intracellular domain (NICD)  
 from the membrane (By similarity).  
 CC -!- PTM: Phosphorylated (By similarity).  
 CC -!- DISEASE: Defects in NOTCH3 are associated with CADASIL (Cerebral  
 autosomal dominant arteriopathy with subcortical infarcts and  
 leukoencephalopathy) which causes a type of stroke and dementia of  
 which key features include recurrent subcortical ischemic events  
 and vascular dementia.  
 CC -!- SIMILARITY: BELONGS TO THE NOTCH FAMILY.  
 CC -!- SIMILARITY: CONTAINS 34 EGF-LIKE DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.  
 CC -!- SIMILARITY: CONTAINS 5 ANK REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U97669; AAB91371.1; -;  
 DR EMBL; AF058900; AAC14346.1; -;  
 DR EMBL; AF058881; AAC14346.1; JOINED.  
 DR EMBL; AF058882; AAC14346.1; JOINED.  
 DR EMBL; AF058883; AAC14346.1; JOINED.  
 DR EMBL; AF058884; AAC14346.1; JOINED.  
 DR EMBL; AF058885; AAC14346.1; JOINED.  
 DR EMBL; AF058886; AAC14346.1; JOINED.  
 DR EMBL; AF058887; AAC14346.1; JOINED.  
 DR EMBL; AF058888; AAC14346.1; JOINED.



STIC-Biotech/ChemLib

98589

From: Slobodyansky, Elizabeth  
Sent: Tuesday, April 01, 2003 3:25 PM  
To: STIC-Biotech/ChemLib  
Subject: 10/023,888

Please search for case 10/023,888:

ES

SEQ ID NOs: 1, 2 and 22 against commercial and interference databases.

Thank you.

Elizabeth Slobodyansky, PhD

Primary Examiner

Art Unit 1652  
CM1 10D11  
703-306-3222

mail box 10D01

Point of Contact:  
Mona Smith  
Technical Information Specialist  
CM1 6A01  
Tel: 308-3278

RECEIVED  
APR - 1 2003  
(STIC)

Searcher: M. Ba, 74  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: 4/9/03  
Date Completed: 4/14/03  
Searcher Prep/Review: 8  
Clerical: \_\_\_\_\_  
Online time: 12

TYPE OF SEARCH:  
NA Sequences: 1  
AA Sequences: 2  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

١٠